1943

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

SECOND SERIES, VOL. 44

PUBLISHED QUARTERLY
AT
MOUNT ROYAL AND GUILFORD AVENUES
BALTIMORE, MD.
BY THE
PRINCETON UNIVERSITY PRESS
PRINCETON, N. J.

ANNALS OF MATHEMATICS
Vol. 44, No. 1, January, 1943

ON HODGE'S THEORY OF HARMONIC INTEGRALS

BY HERMANN WEYL

(Received June 4, 1942)

The attempt which W. V. D. Hodge made in Chapter III of his beautiful book[1] to establish *the existence of harmonic integrals with preassigned periods* has not been entirely successful because the proof is partly based on a false statement (p. 136) concerning the behavior of the solution of a non-homogeneous integral equation when the spectrum parameter approaches an eigen value. In a Princeton seminar on the subject, H. F. Bohnenblust pointed out that counter examples are readily available even for linear equations with a finite number of unknowns. For instance the equation $\lambda x + Ax = c$ with

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad c = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

is solvable for $\lambda = 0$ (x_1 arbitrary, $x_2 = 1$) and yet the solution for $\lambda \neq 0$,

$$x_1 = 1/\lambda, \qquad x_2 = 0$$

does not tend to a limit with $\lambda \to 0$.

In his book Hodge uses the *parametrix method* first developed for a single elliptic differential equation by E. E. Levi and D. Hilbert.[2] Building on the formal foundations laid by Hodge, I will show here how the argument can be made conclusive. Hilbert's procedure served me as a model.

Let n be the dimensionality of our Riemannian manifold. I denote by $^*u$, Du the dual form and the derivative of any (linear differential) form u and use the abbreviation Δ for the operator D^*D.[3] For two forms u, v of rank p, $n - p$ respectively (v, u) designates the integral of the product $v \cdot u$ over the whole manifold. $(^*u, u)$ is positive unless $u = 0$. An immediate consequence is

LEMMA I. *$\Delta u = 0$ implies $Du = 0$.*

Indeed $(D^*Du, u) = 0$ leads to $(^*Du, Du) = 0$, hence $Du = 0$.

In the following, f, u, φ, η are forms of rank p and g, v, ψ, ϑ forms of rank $n - p$.

---

[1] *The Theory and Applications of Harmonic Integrals*, Cambridge, 1941. See also Proc. London Math. Soc. (2) **41**, 1936, pp. 483–496 where Hodge ascribes the idea of using Hilbert's parametrix method to H. Kneser. I find it hard to judge whether a previous proof along different lines (Proc. London Math. Soc. (2) **38**, 1933, p. 72) is complete, or rather how much effort is needed to make it complete. For the Euclidean case, see W. V. D. Hodge, Proc. London Math. Soc. (2) **36**, 1932, p. 257, and H. Weyl, Duke Math. Jour. **7**, 1940, pp. 411–444.

[2] E. E. Levi, Memorie della Società italiana delle Scienze, Ser. 3ª, Tom. 16, 1909; D. Hilbert, *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen*, Leipzig, 1912, pp. 223–231.

[3] The author intended $*u$, the printer evidently disliked it and replaced it by $^*u$. If a standardized notation in the theory of linear differential forms is adopted the author would recommend to follow him and not the printer.

The rank p is fixed; no induction with respect to p takes place. The goal is to prove the following

THEOREM I. *For any given null form g, $g \sim 0$, the equation*

$$\Delta u = g \tag{1}$$

has a solution u.

I copy Hodge's two basic formulas (3) and (4) on pp. 132, 133 of his book, replacing $p - 1$ by p and using the abbreviation $1/\gamma = (-1)^{np}(n - 2)\alpha_n$. Let K be the operator with the kernel $\gamma \cdot K_p(x, y)$ which carries any form $u(x)$ into $\gamma \int K_p(x, y) \cdot u(y)$, and K' its transpose. The "parametrix" operators Q, P with the kernels $\gamma \cdot \omega_p(x, y)$ and $\gamma \cdot \omega_{p-1}(x, y)$ are symmetric,

$$(Qv, g) = (Qg, v).$$

Finally, I set $DPD = \Pi$. Hodge's formulas read

$$Ku - u = Q\Delta u + (-1)^n \Pi^* u, \tag{2}$$

$$K'v - v = \Delta Qv + (-1)^{n*} \Pi v. \tag{2'}$$

The solutions of the equations

$$Ku - u = 0, \qquad K'v - v = 0$$

will be called the eigen forms of the kernels K and K' (*scilicet* "for the eigen value 1"). We try to solve our problem by means of the non-homogeneous integral equation suggested by (2),

$$Ku - u = Qg. \tag{E}$$

It is essential to study this equation not only for null forms but in a wider set $\mathfrak{G}$; the success of the method depends on the proper choice of that linear space $\mathfrak{G}$. Here is my definition:

g belongs to $\mathfrak{G}$ whenever PDg is closed,

$$DPDg = \Pi g = 0.$$

Every form of the type

$$f = \Pi v \qquad (v \text{ arbitrary})$$

is said to belong to $\mathfrak{F}$. Evidently $\mathfrak{G}$ contains all *closed* forms g whereas all elements f of $\mathfrak{F}$ are *null* forms. $\mathfrak{F}$ and $\mathfrak{G}$ are orthogonal:

LEMMA II. *$(g, f) = 0$ for $g \in \mathfrak{G}$, $f \in \mathfrak{F}$.*

Indeed, if PDg is closed, then

$$(PDg, Dv) = 0 = (Dg, PDv),$$

an equation which may at once be changed into

$$(g, DPDv) = 0.$$

I take over Hodge's Lemma I on p. 142:

LEMMA III. *If ψ is any eigen form of K' then $Q\psi$ is closed.*

For the sake of completeness I repeat the simple proof. Equation (2′) yields for $\xi = Q\psi$:

$$\Delta\xi = (-1)^{n-1}{}^*\Pi\psi, \tag{3}$$

hence $D^*\Delta\xi = D^*D^*D\xi = 0$ and then by double application of Lemma I,[4]

$$D^*D\xi = 0, \qquad D\xi = 0.$$

Incidentally we learn from (3) and the intermediate equation $\Delta\xi = 0$ that $\Pi\psi = 0$, or that *the eigen forms ψ of K' lie in $\mathfrak{G}$.*

We analyze the eigen forms of K and K' as follows. Within the linear space of all eigen forms $\bar{\varphi}$ of K we consider the subspace $\mathfrak{f}$ of the *closed* eigen forms φ and choose our basis

$$\varphi_1, \cdots, \varphi_l, \qquad \bar{\varphi}_1, \cdots, \bar{\varphi}_m$$

for all eigen forms accordingly, i.e. $\varphi_1, \cdots, \varphi_l$ span $\mathfrak{f}$. Equation (2) yields

$$Q\Delta\bar{\varphi} = (-1)^{n-1}\Pi^*\bar{\varphi}. \tag{4}$$

This proves on the one hand that each closed eigen form φ of K satisfies the condition $\Pi^*\varphi = 0$,

LEMMA IV. *${}^*\varphi \in \mathfrak{G}$ for every $\varphi \in \mathfrak{f}$.*

It shows on the other hand that $\bar{\psi} = \Delta\bar{\varphi}$ satisfies the conditions

$$\Delta Q\bar{\psi} = 0, \qquad \Pi\bar{\psi} = 0$$

because the operators $\Delta\Pi$ and $\Pi\Delta$ annihilate. It then follows from (2′) that $\bar{\psi}$ is an eigen form of K'. The m forms $D\bar{\varphi}_1, \cdots, D\bar{\varphi}_m$ are linearly independent by construction, and hence by Lemma I the same is true for the forms

$$\bar{\psi}_1 = \Delta\bar{\varphi}_1, \cdots, \qquad \bar{\psi}_m = \Delta\bar{\varphi}_m.$$

The transposed kernel K' has the same number $l + m$ of linearly independent eigen forms as K. We determine a basis

$$\bar{\psi}_1, \cdots, \bar{\psi}_m; \qquad \psi_1, \cdots, \psi_l \tag{5}$$

of which the $\bar{\psi}$'s are a part.

The integral equation (E) is solvable if and only if

$$(Qg, \psi) = 0 = (g, Q\psi)$$

for every eigen form ψ of K', or with the notation $\xi = Q\psi$, if

$$(g, \xi) = 0. \tag{6}$$

[4] One differentiation may be saved here by applying the formula $(Ds, Dt) = 0$ holding for any two forms s, t with continuous first derivatives of rank $p - 1$ and $n - p - 1$ (see Weyl, l.c.[1], p. 426) to $s = PD\psi$ and $t = {}^*D\xi$ with the result $(\Pi\psi, \Delta\xi) = 0 = ({}^*\Delta\xi, \Delta\xi)$ whence $\Delta\xi = 0 = \Pi\psi$.

Let us say that ψ is *of the first kind* when $\xi = Q\psi \in \mathfrak{F}$. The forms $\bar{\psi}_1, \cdots, \bar{\psi}_m$ are of the first kind, on account of the equation (4). We choose our basis (5) so that

$$\bar{\psi}_1, \cdots, \bar{\psi}_m; \qquad \psi_1, \cdots, \psi_\nu$$

span the linear manifold of all eigen forms of K' of the first kind. By Lemma II the relation (6) holds good for *any* $g \in \mathfrak{G}$ in case ψ is of the first kind, and thus the $m + l$ conditions (6) reduce to the last $l - \nu$ of them,

$$(g, Q\psi_{\nu+1}) = 0, \cdots, \qquad (g, Q\psi_l) = 0. \tag{7}$$

Let $\mathfrak{G}_1$ denote the set of those forms $g \in \mathfrak{G}$ which satisfy the conditions (7). *We have found that under the assumption $g \in \mathfrak{G}_1$ the integral equation* (E) *has a solution* u.

For this solution u we obtain from (2):

$$Q(g - \Delta u) = (-1)^n \cdot \Pi^* u, \tag{8}$$

hence $\Delta Q(g - \Delta u) = 0$. Combining this with $\Pi(g - \Delta u) = 0$ and applying (2′) to $v = g - \Delta u$ one finds

$$g - \Delta u = \psi = \bar{c}_1\bar{\psi}_1 + \cdots + \bar{c}_m\bar{\psi}_m + c_1\psi_1 + \cdots + c_l\psi_l$$

to be an eigen form of K'. More precisely, because of (8), $Q\psi \in \mathfrak{F}$, ψ is an eigen form of the first kind, which forces $c_{\nu+1}, \cdots, c_l$ to vanish. Writing u for $u + \bar{c}_1\bar{\varphi}_1 + \cdots + \bar{c}_m\bar{\varphi}_m$ we arrive at the following

INTERMEDIARY PROPOSITION: *For any $g \in \mathfrak{G}_1$ there exists a form u and ν constants $c_1, \cdots, c_\nu$ such that*

$$g - \Delta u = c_1\psi_1 + \cdots + c_\nu\psi_\nu. \tag{9}$$

We know from Lemma IV that the dual form $^*\varphi$ of any element φ of $\mathfrak{f}$ lies in $\mathfrak{G}$. That subspace of $\mathfrak{f}$ the elements φ of which satisfy the conditions

$$(^*\varphi, Q\psi_{\nu+1}) = 0, \quad \cdots, \quad (^*\varphi, Q\psi_l) = 0$$

is of a dimensionality $\mu \geqq \nu$. Let the basis $\varphi_1, \cdots, \varphi_l$ of $\mathfrak{f}$ be so chosen that $\varphi_1, \cdots, \varphi_\mu$ span this subspace. From (9) we obtain for the ν unknowns c_β the μ linear equations

$$\sum_\beta H_{\alpha\beta} \cdot c_\beta = (g, \varphi_\alpha) \qquad \begin{pmatrix} \alpha = 1, \cdots, \mu; \\ \beta = 1, \cdots, \nu \end{pmatrix} \tag{10}$$

where

$$H_{\alpha\beta} = (\psi_\beta, \varphi_\alpha).$$

I maintain:

LEMMA V. *$\| H_{\alpha\beta} \|$ is a non-singular square matrix.*

Once this is established we have reached the goal. For then the ν conditions

$$(g, \varphi_\alpha) = 0 \qquad (\alpha = 1, \cdots, \nu)$$

imply $c_\alpha = 0$ whereby (9) reduces to $g - \Delta u = 0$. In other words, if $g \in \mathfrak{G}$ satisfies the relations

$$(11)\quad (g, Q\psi_{\nu+1}) = 0, \cdots, (g, Q\psi_l) = 0; \qquad (g, \varphi_1) = 0, \cdots, (g, \varphi_\nu) = 0$$

then the equation (1) is solvable. A null form g fulfills all our requirements, because the φ_i and $Q\psi_i$ are closed, the first by construction, the others by Lemma III.

Proof of Lemma v. We have found the equations (10) to be solvable if $g \in \mathfrak{G}_1$. For

$$(12)\qquad \varphi = a_1\varphi_1 + \cdots + a_\mu\varphi_\mu$$

the integral $(^*\varphi, \varphi)$ is a positive definite quadratic form of $a_1, \cdots, a_\mu$. Hence we can determine the coefficients a_i in (12) so as to assign arbitrary values b_α to the integrals

$$(^*\varphi, \varphi_\alpha) \qquad (\alpha = 1, \cdots, \mu).$$

But $g = {}^*\varphi \in \mathfrak{G}_1$. Hence we see that the equations

$$\textstyle\sum_\beta H_{\alpha\beta}c_\beta = b_\alpha \qquad \begin{pmatrix}\alpha = 1, \cdots, \mu; \\ \beta = 1, \cdots, \nu\end{pmatrix}$$

have a solution c_β for arbitrary b_α. In view of $\mu \geqq \nu$ this statement is equivalent to our lemma.

In proving Theorem I we actually showed that the equation $\Delta u = g$ is solvable if $g \in \mathfrak{G}$ satisfies the conditions (11). Hence each such g is a null form, and the linear space $\mathfrak{G}$ is of finite dimensionality $\leqq l$ modulo the space of null forms. As $\mathfrak{G}$ contains all closed forms of rank $n - p$, we find *a fortiori* that the number R'_{n-p} of linearly independent closed forms of rank $n - p$ modulo null is finite and $\leqq l$. The conditions (11) are of the type $(g, f) = 0$ where f runs over certain specified closed forms of rank p. Consider the "inner product" (g, f) of any two closed forms g, f of rank $n - p$ and p respectively; the factors matter only modulo null. Our proof implies this further fact:

Theorem ii. *If the inner product (g, f) vanishes for a given closed g and all closed f, then $g \sim 0$.*

It is of course also true that the product cannot vanish for a given closed f and all closed g unless $f \sim 0$. Both facts together give the duality law

$$(13)\qquad R'_{n-p} = R'_p .$$

Theorem II has nothing to do with any Riemannian metric. de Rham's second theorem follows at once from it by means of the expression of the product (g, f) in terms of the periods of g and f (Hodge, p. 85, last line), but it is essentially simpler since it deals with closed forms only, and not with forms and cycles. Its proof on an arbitrary manifold should be correspondingly easier.

The following proposition is equivalent to Theorem I for the rank $p - 1$ instead of p:

THEOREM III. *For any form f there exists a uniquely determined $\eta \sim f$ such that $^*\eta$ is closed. If f be closed, then η is harmonic.*

Indeed, set $f = Dt + \eta$, t being of rank $p - 1$. The requirement $D^*\eta = 0$ leads to the equation $D^*Dt = D^*f$ which is solvable by Theorem I.

The new proposition shows at once that for any rank p the space of closed forms modulo null may be identified with the space of harmonic forms. This makes the equation (13) particularly lucid because $^*u$ is harmonic if u is and vice versa. The same proposition provides another proof for Theorem II, because one has merely to substitute $^*\eta$ for ϑ in order to see that the vanishing of the inner product (η, ϑ) of a fixed harmonic form η with every harmonic ϑ implies $\eta = 0$. The observation (Hodge, p. 139) that on account of (2) the harmonic p-forms are eigen forms of K again proves the inequality $R'_p \leqq l$.

The link with the homology theory of cycles is established by de Rham's first theorem stating that a p-cycle C is homologous zero if the integral of every closed p-form f over C vanishes.

INSTITUTE FOR ADVANCED STUDY, PRINCETON, N. J.

ANNALS OF MATHEMATICS
Vol. 44, No. 1, January, 1943

MOUTARD-ČECH HYPERQUADRICS ASSOCIATED WITH A POINT OF A HYPERSURFACE

BY BUCHIN-SU

(Received May 22, 1942)

The purpose of this paper is to give a new proof of the theorem of Čech[1] concerning the locus of the quadrics of Čech of sections produced by certain spaces and to generalize, as a consequence of the equation to a Moutard-Čech hyperquadric, the notion of Moutard pencils[2] to the theory of hypersurfaces.

Let O be a generic point of a hypersurface V_n in a projective space S_{n+1}. If O is taken for the origin of coordinates with the tangent hyperplane of V_n at O in non-homogeneous coordinates $X^1, \cdots, X^{n+1}$ as the coordinate hyperplane $X^{n+1} = 0$, then the expansion of V_n at O is of the form

$$X^{n+1} = \tfrac{1}{2}H_{\sigma\tau}X^\sigma X^\tau + \tfrac{1}{3}K_{\sigma\tau\rho}X^\sigma X^\tau X^\rho + \tfrac{1}{12}H_{\sigma\tau\rho u}X^\sigma X^\tau X^\rho X^u + \cdots, \tag{1}$$

where we have made use of the convention that when the same index appears as a subscript and superscript in a term this term stands for the sum of the terms obtained by giving the index each of its n values $1, 2, \cdots, n$. Without loss of generality we may further assume that the coefficients in (1) are symmetrical in the subscripts.

We shall calculate the fundamental quantities $a_{ij}(i, j = 1, \cdots, n)$ of the fundamental form F_2 of Fubini.[3] Putting

$$b_{\sigma\tau}\, dX^\sigma dX^\tau = \left(X \frac{\partial X}{\partial X^1} \cdots \frac{\partial X}{\partial X^n}\, d^2X\right)$$

for homogeneous coordinates with $X^0 = 1$, and noticing that

$$\frac{\partial X^i}{\partial X^j} = \delta^i_j \qquad (i, j = 1, \cdots, n),$$

we have

$$b_{rs} = \frac{\partial^2 X^{n+1}}{\partial X^r \partial X^s} = H_{rs} + 2K_{rs\rho}X^\rho + H_{rs\rho\tau}X^\rho X^\tau + (3), \tag{2}$$

where (m) denotes all of the terms of order $\geqq m$ in $X^1, \cdots, X^n$. The discriminant B of the above form is consequently found to be

$$B = H\{1 + 2H^{\sigma\tau}K_{\sigma\tau\rho}X^\rho + L_{\sigma\tau}X^\sigma X^\tau + (3)\}, \tag{3}$$

---

[1] Cf. G. Fubini and E. Čech, *Geometria proiettiva differenziale*, vol. II, Bologna (1927), p. 618; in the sequel referred to as *G. P. D.*

[2] For the definition of a Moutard pencil in S_3 see B. Su and A. Ichida, *On certain cones connected with a surface in the affine space*, Japanese Journal of Mathematics, 10 (1933), pp. 209–216.

[3] Cf. *G. P. D.*, p. 608.

where

(4) $$L_{vw} = H^{\sigma\tau}H_{\sigma\tau vw} + \tfrac{1}{2}H^{\sigma u,\tau\rho}(K_{\sigma\tau v}K_{u\rho w} - K_{u\tau v}K_{\sigma\rho w} + K_{\sigma\tau w}K_{u\rho v} - K_{u\tau w}K_{\sigma\rho v})$$

and H is the determinant of H_{rs}, H^{ij} is the algebraic complementary of H_{ij} in H, divided by H and, finally, $H^{ij,kl} = 0$ if $i = j$ or $k = l$, and otherwise $H^{ij,kl}$ is the minor formed by striking out the rows and columns in H which contain H_{ik}, H_{il}, H_{jk}, H_{jl}, also divided by H, the sign being $+1$ or -1 according as the permutations i, j and the remaining rows in order, and k, l and the remaining columns in the order $1, 2, \cdots, n$ are of the same or opposite parity.

From (3) it follows that

(5) $$\begin{aligned} B^{-1/(n+2)} = H^{-1/(n+2)} \Big\{ 1 &- \frac{2}{n+2} H^{\sigma\tau}K_{\sigma\tau\rho}X^\rho \\ &+ \left(\frac{2(n+3)}{(n+2)^2} H^{\sigma\tau}H^{\rho u}K_{\sigma\tau v}K_{\rho uw} - \frac{1}{n+2}L_{vw}\right)X^vX^w + (3)\Big\}, \end{aligned}$$

so that

(6) $$\begin{aligned} a_{rs} &= B^{-1/(n+2)}b_{rs} \\ &= H^{-1/(n+2)}\left\{H_{rs} + 2\left(K_{rs\rho} - \frac{1}{n+2}H_{rs}H^{\sigma\tau}K_{\sigma\tau\rho}\right)X^\rho + (2)\right\}. \end{aligned}$$

Thus we obtain *at O*

(7) $$\left\{\begin{aligned} a_{rs} &= H^{-1/(n+2)}H_{rs} && (r, s = 1, \cdots, n), \\ \frac{\partial a_{rs}}{\partial X^t} &= 2H^{-1/(n+2)}\left(K_{rst} - \frac{1}{n+2}H_{rs}H^{\sigma\tau}K_{\sigma\tau t}\right) && (r, s, t, = 1, \cdots, n), \\ a^{rs} &= H^{1/(n+2)}H^{rs}, \\ \Gamma_{rs,t} &= \frac{1}{2}\left(\frac{\partial a_{rt}}{\partial X^s} + \frac{\partial a_{st}}{\partial X^r} - \frac{\partial a_{rs}}{\partial X^t}\right) \\ &= H^{-1/(n+2)}\Big\{K_{rst} - \frac{1}{n+2}H_{rt}H^{\sigma\tau}K_{\sigma\tau s} \\ &\qquad - \frac{1}{n+2}H_{st}H^{\sigma\tau}K_{\sigma\tau r} + \frac{1}{n+2}H_{rs}H^{\sigma\tau}K_{\sigma\tau t}\Big\}, \\ \Gamma^l_{rs} &= H^{l\sigma}K_{rs\sigma} - \frac{1}{n+2}H^{\sigma\tau}K_{\sigma\tau s}\delta^l_r \\ &\qquad - \frac{1}{n+2}H^{\sigma\tau}K_{\sigma\tau r}\delta^l_s + \frac{1}{n+2}H^{l\rho}H_{rs}H^{\sigma\tau}K_{\sigma\tau\rho}. \end{aligned}\right.$$

The covariant derivatives of X^i are

$$X^i_{rs} = \frac{\partial^2 X^i}{\partial X^r \partial X^s} - \Gamma^l_{rs}\frac{\partial X^i}{\partial X^l} \qquad (i = 0, 1, \cdots, n+1).$$

At O we obtained from (7), X^0 being 1,

$$\left\{\begin{aligned} X^0_{rs} &= 0, \\ X^i_{rs} &= -\Gamma^i_{rs} \qquad\qquad (i = 1, \cdots, n), \\ X^{n+1}_{rs} &= H_{rs}. \end{aligned}\right. \tag{8}$$

Introduce, as usual, the vector $\mathfrak{X}$ of components[4]

$$\mathfrak{X}^i = \frac{1}{n}\Delta_2 X^i = \frac{1}{n} a^{\sigma\tau} X^i_{\sigma\tau} \qquad (i = 0, 1, \cdots, n+1) \tag{9}$$

and calculate the values at O by means of (7) and (8). The result of this computation is as follows:

$$\left\{\begin{aligned} \mathfrak{X}^0 &= 0, \\ \mathfrak{X}^i &= -\frac{2}{n+2} H^{1/(n+2)} H^{\sigma\tau} K_{\sigma\tau\rho} H^{i\rho} \quad (i = 1, \cdots, n), \\ \mathfrak{X}^{n+1} &= H^{1/(n+2)}. \end{aligned}\right. \tag{10}$$

We are in a position to consider the corresponding tangential coordinates $\xi^i (i = 0, 1, \cdots, n+1)$ of Čech given by the minors of order n in the following matrix[5]:

$$\left\| \begin{matrix} 1 & X^1 & X^2 & \cdots & X^n & X^{n+1} \\ 0 & 1 & 0 & \cdots & 0 & \dfrac{\partial X^{n+1}}{\partial X^1} \\ 0 & 0 & 1 & \cdots & 0 & \dfrac{\partial X^{n+1}}{\partial X^2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & \dfrac{\partial X^{n+1}}{\partial X^n} \end{matrix} \right\| : B^{1/(n+2)},$$

or, more precisely,

$$\left\{\begin{aligned} \xi^i &= B^{-1/(n+2)} \frac{\partial X^{n+1}}{\partial X^i} \qquad (i = 1, \cdots, n), \\ \xi^{n+1} &= B^{-1/(n+2)}. \end{aligned}\right. \tag{11}$$

A reference to (1) and (5) gives

$$\xi^i = H^{-1/(n+2)} \left\{ H_{i\rho} X^\rho + \left(K_{ivw} - \frac{2}{n+2} H_{iv} H^{\sigma\tau} K_{\sigma\tau w} \right) X^v X^w + (3) \right\} \tag{12}$$

$$(i = 1, \cdots, n).$$

---

[4] Cf. *G. P. D.*, p. 611. For clearness of notation the X there utilized is now replaced by $\mathfrak{X}$.

[5] Cf. *G. P. D.*, p. 609.

so that at O

$$\frac{\partial \xi^i}{\partial X^l} = H^{-1/(n+2)} H_{il} \qquad (i, l = 1, \cdots, n),$$

$$\frac{\partial^2 \xi^i}{\partial X^r \partial X^s} = 2H^{-1/(n+2)} \left(K_{irs} - \frac{1}{n+2} H_{ir} H^{\sigma\tau} K_{\sigma\tau s} - \frac{1}{n+2} H_{is} H^{\sigma\tau} K_{\sigma\tau r} \right) \qquad (r, s = 1, \cdots, n)$$

and therefore

$$\text{(13)} \quad \xi^i_{rs} = H^{-1/(n+2)} \left\{ K_{irs} - \frac{1}{n+2} H^{\sigma\tau} K_{\sigma\tau s} H_{ir} - \frac{1}{n+2} H^{\sigma\tau} K_{\sigma\tau r} H_{is} - \frac{1}{n+2} H^{\sigma\tau} K_{\sigma\tau i} H_{rs} \right\} \qquad (i = 1, \cdots, n).$$

If we define, as the dual of $\mathfrak{X}$, the vector Ξ of components

$$\text{(14)} \quad \Xi^j = \frac{1}{n} \Delta_2 \xi^j = \frac{1}{n} a^{rs} \xi^j_{rs} \qquad (j = 0, 1, \cdots, n+1),$$

we have at 0

$$\Xi^i = 0 \qquad (i = 1, \cdots, n),$$

$$\text{(15)} \quad \Xi^{n+1} = \frac{4}{n(n+2)} \left\{ \frac{2n+3}{n+2} H^{\sigma\tau} H^{\rho u} H^{rs} K_{\sigma\tau r} K_{\rho u s} - \tfrac{1}{2} H^{rs} L_{rs} \right\},$$

whence

$$\text{(16)} \quad S\mathfrak{X}\Xi = \frac{4}{n(n+2)} H^{1/(n+2)} \left\{ \frac{2n+3}{n+2} H^{\sigma\tau} H^{\rho u} H^{rs} K_{\sigma\tau r} K_{\rho u s} - \tfrac{1}{2} H^{rs} L_{rs} \right\}.$$

The quadric of Čech of V_n at O is given by the equation[6]

$$\text{(17)} \quad 2\lambda\mu + (\mu)^2 S\mathfrak{X}\Xi - a_{\sigma\tau} \nu^\sigma \nu^\tau = 0$$

where the local coordinates $\lambda, \nu^1, \cdots, \nu^n, \mu$ of a point are related to the coordinates $X^1, \cdots, X^{n+1}$ by the following equations:

$$\text{(18)} \quad 1 : X^1 : X^2 : \cdots : X^n : X^{n+1} = \lambda : \mu\mathfrak{X}^1 + \nu^1 : \mu\mathfrak{X}^2 + \nu^2 : \cdots : \mu\mathfrak{X}^n + \nu^n : \mu\mathfrak{X}^{n+1}.$$

Substituting (10) and (16) into (17) and reducing by (18), we have the equation to the quadric of Čech of V_n at O:

$$\text{(19)} \quad X^{n+1} - \tfrac{1}{2} H_{\sigma\tau} X^\sigma X^\tau - \frac{2}{n+2} H^{\rho u} K_{\rho u \sigma} X^\sigma X^{n+1} + \left\{ \frac{2(n+3)}{n(n+2)^2} H^{\sigma\tau} H^{\rho u} H^{rs} K_{\sigma\tau r} K_{\rho u s} - \frac{1}{n(n+2)} H^{rs} L_{rs} \right\} (X^{n+1})^2 = 0.$$

[6] Cf. *G. P. D.*, p. 616.

From the last equation we can easily demonstrate some known results. In fact, consider a hyperquadric Q which has at O a contact of the second order with V_n. Since the equation of such a quadric Q is expressible in the form

$$X^{n+1} = \tfrac{1}{2}H_{\sigma\tau}X^\sigma X^\tau + a_\rho X^\rho X^{n+1} + a(X^{n+1})^2, \tag{20}$$

the quadric of Čech given by (19) evidently belongs to this family. It is easily seen that the tangents to the intersection of V_n and Q drawn from O constitute a hypercone Γ^3_{n-1} of the third order

$$F_3 \equiv (\tfrac{1}{3}K_{\sigma\tau\rho} - \tfrac{1}{2}a_\rho H_{\sigma\tau})X^\sigma X^\tau X^\rho = 0. \tag{21}$$

The latter is apolar to the asymptotic hypercone

$$H_{\sigma\tau}X^\sigma X^\tau = 0 \tag{22}$$

if and only if

$$H^{\sigma\tau}(2K_{\sigma\tau i} - a_i H_{\sigma\tau} - 2a_\tau H_{\sigma i}) = 0 \qquad (i = i, \cdots, n),$$

namely,

$$a_i = \frac{2}{n+2} H^{\sigma\tau} K_{\sigma\tau i} \qquad (i = 1, \cdots, n). \tag{23}$$

In consequence, the quadric Q in consideration must belong to the pencil

$$X^{n+1} = \tfrac{1}{2}H_{\sigma\tau}X^\sigma X^\tau + \frac{2}{n+2} H^{\rho u} K_{\rho u \sigma} X^\sigma X^{n+1} + a(X^{n+1})^2, \tag{24}$$

called the Darboux pencil.[7] The equation (19) shows that the quadric of Čech is contained in the Darboux pencil.[8]

We now proceed to find the section V_ν of V_n produced by a space $[\nu + 1]$ of $\nu + 1$ dimensions through a given tangent space $[\nu]$ of ν dimensions at O. First it is convenient to take the given space $[\nu]$ through O for the space

$$X^{\nu+1} = 0, \cdots, X^{n+1} = 0 \qquad (\nu \geqq 1), \tag{25}$$

so that the $[\nu + 1]$ is then given by the equations

$$X^k = \lambda^k X^{n+1} \qquad (k = \nu + 1, \cdots, n). \tag{26}$$

In virtue of (1) and (26) there is no difficulty in showing that the section of V_n produced by (26) is a hypersurface V_ν in this $[\nu + 1]$, the expansion of V_ν at O being

$$\begin{aligned} X^{n+1} &= \frac{1}{2}\sum_1^\nu H_{\sigma\tau}X^\sigma X^\tau + \frac{1}{3}\sum_1^\nu \bar{K}_{\sigma\tau\rho}X^\sigma X^\tau X^\rho \\ &\qquad + \frac{1}{12}\sum_1^\nu \bar{H}_{\sigma\tau\rho u}X^\sigma X^\tau X^\rho X^u + \cdots, \end{aligned} \tag{27}$$

[7] J. Kanitani, *Géométrie différentielle projective des hypersurfaces*, Ryojun (1931), p. 33.
[8] Cf. *G. P. D.*, p. 617.

where the coefficients symmetrical in the subscripts are given by the following equations:

$$
(28)\quad \left\{
\begin{aligned}
\bar{K}_{\sigma\tau\rho} &= K_{\sigma\tau\rho} + \tfrac{1}{2}H_{\tau\rho}\sum_{\alpha=\nu+1}^{n} H_{\sigma\alpha}\lambda^{\alpha} \\
&\qquad + \tfrac{1}{2}H_{\rho\sigma}\sum_{\alpha=\nu+1}^{n} H_{\tau\alpha}\lambda^{\alpha} + \tfrac{1}{2}H_{\sigma\tau}\sum_{\alpha=\nu+1}^{n} H_{\rho\alpha}\lambda^{\alpha}, \\
\bar{H}_{\sigma\tau\rho u} &= H_{\sigma\tau\rho u} + K_{\tau\rho u}\sum_{\alpha=\nu+1}^{n} H_{\sigma\alpha}\lambda^{\alpha} + K_{\rho u\sigma}\sum_{\alpha=\nu+1}^{n} H_{\tau\alpha}\lambda^{\alpha} + K_{u\sigma\tau}\sum_{\alpha=\nu+1}^{n} H_{\rho\alpha}\lambda^{\alpha} \\
&\quad + K_{\sigma\tau\rho}\sum_{\alpha=\nu+1}^{n} H_{u\alpha}\lambda^{\alpha} + \sum_{\alpha,\beta=\nu+1}^{n} (H_{\sigma\tau}H_{\rho\alpha}H_{u\beta} + H_{\sigma\rho}H_{u\alpha}H_{\tau\beta} \\
&\quad + H_{\sigma u}H_{\tau\alpha}H_{\rho\beta} + H_{\tau\rho}H_{u\alpha}H_{\sigma\beta} + H_{\tau u}H_{\sigma\alpha}H_{\rho\beta} + H_{\rho u}H_{\sigma\alpha}H_{\tau\beta})\lambda^{\alpha}\lambda^{\beta} \\
&\quad + \sum_{\alpha=\nu+1}^{n} (K_{\sigma\tau\alpha}H_{\rho u} + K_{\sigma\rho\alpha}H_{u\tau} + K_{\sigma u\alpha}H_{\tau\rho} + K_{\tau\rho\alpha}H_{\sigma u} + K_{\tau u\alpha}H_{\sigma\rho} \\
&\quad + K_{\rho u\alpha}H_{\sigma\tau})\lambda^{\alpha} + \tfrac{1}{2}(H_{\sigma\tau}H_{\rho u} + H_{\sigma\rho}H_{\tau u} + H_{\rho\tau}H_{\sigma u}) \\
&\qquad \cdot \sum_{\alpha,\beta=\nu+1}^{n} H_{\alpha\beta}\lambda^{\alpha}\lambda^{\beta} \qquad (\sigma,\tau,\rho,u = 1,\cdots,\nu).
\end{aligned}
\right.
$$

In order to find the quadric of Čech of V_ν at O we have only to replace in (19) n by ν; K_{ijk}, H_{ijkl}, etc. by $\bar{K}_{ijk}$, $\bar{H}_{ijkl}$, etc., remembering that in this case the summation must be of the terms obtained by giving the index each of its ν values $1, 2, \cdots, \nu$. Thus we have

$$
(29)\quad
\begin{aligned}
&X^{n+1} - \frac{1}{2}\sum_{\sigma,\tau=1}^{\nu} H_{\sigma\tau}X^{\sigma}X^{\tau} - \frac{2}{\nu+2}\sum_{\rho,\sigma,u=1}^{\nu} \bar{H}^{\rho u}\bar{K}_{\rho u\sigma}X^{\sigma}X^{n+1} \\
&+\left\{\frac{2(\nu+3)}{\nu(\nu+2)^2}\sum_{1}^{\nu} \bar{H}^{\sigma\tau}\bar{H}^{\rho u}\bar{K}_{\sigma\tau r}\bar{K}_{\rho u s}\bar{H}^{rs} - \frac{1}{\nu(\nu+2)}\sum_{1}^{\nu} \bar{H}^{rs}\bar{L}_{rs}\right\}(X^{n+1})^2 = 0,
\end{aligned}
$$

where

$$
(30)\quad
\begin{aligned}
\bar{L}_{rs} = \sum_{1}^{\nu} \bar{H}^{\sigma\tau}\bar{H}_{\sigma\tau rs} + \frac{1}{2}\sum_{1}^{\nu} \bar{H}^{\sigma u,\tau\rho}(\bar{K}_{\sigma\tau r}\bar{K}_{\rho u s} - \bar{K}_{u\tau r}\bar{K}_{\sigma\rho s} \\
+ \bar{K}_{\sigma\tau s}\bar{K}_{\rho u r} - \bar{K}_{u\tau s}\bar{K}_{\sigma\rho r}).
\end{aligned}
$$

From (28) it follows that

$$
\sum_{1}^{\nu} \bar{H}^{\rho u}K_{\rho u\sigma} = \sum_{1}^{\nu} \bar{H}^{\rho u}K_{\rho u\sigma} + \tfrac{1}{2}(\nu+2)\sum_{\alpha=\nu+1}^{n} H_{\sigma\alpha}\lambda^{\alpha} \qquad (\sigma = 1,\cdots,\nu).
$$

Hence, putting

$$
(31)\quad \left\{
\begin{aligned}
\mathfrak{A} &= \sum_{1}^{\nu} \bar{H}^{\sigma\tau}\bar{H}^{\rho u}\bar{H}^{rs}\bar{K}_{\sigma\tau r}\bar{K}_{\rho u s}, \\
\mathfrak{B} &= \sum_{1}^{\nu} \bar{H}^{\sigma\tau}\bar{H}^{rs}\bar{H}_{\sigma\tau rs}, \\
\mathfrak{C} &= \sum_{1}^{\nu} \bar{H}^{\sigma u,\tau\rho}\bar{H}^{rs}(\bar{K}_{\sigma\tau r}\bar{K}_{\rho u s} - \bar{K}_{u\tau r}\bar{K}_{\rho\sigma s}),
\end{aligned}
\right.
$$

the equation of the quadric of Čech in consideration can be written in the form

$$
(32)\quad X^{n+1} - \frac{1}{2}\sum_1^\nu H_{\sigma\tau}X^\sigma X^\tau - \left\{\frac{2}{\nu+2}\sum_1^\nu \bar{H}^{\rho u}K_{\rho u\sigma}X^\sigma + \sum_{\sigma=1}^{\nu}\sum_{\alpha=\nu+1}^{n} H_{\sigma\alpha}\lambda^\alpha X^\sigma\right\}X^{n+1} + \left\{\frac{2(\nu+3)}{\nu(\nu+2)^2}\mathfrak{A} - \frac{1}{\nu(\nu+2)}(\mathfrak{B}+\mathfrak{C})\right\}(X^{n+1})^2 = 0.
$$

Since $\bar{K}_{\sigma\tau\rho}$ and $\bar{H}_{\sigma\tau\rho u}$ are respectively linear and quadratic in λ's and $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$ given by (31) are all quadratic in these parameters, by eliminating them from (26) and (32) we may easily conclude that when the space $[\nu+1]$ turns about the fixed space $[\nu]$ given by (25) the quadric of Čech (32) describes a hyperquadric, which completes the proof of the theorem of Čech.

For the subsequent development it is, however, desirable to derive the explicit form for $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$ and therefore the equation of the hyperquadric thus obtained, namely, the Moutard-Čech hyperquadric belonging to the given space $[\nu]$. The first two of them may easily be calculated by means of (28). Thus we obtain

$$
(33)\quad \begin{aligned}
\mathfrak{A} = {}& \sum_1^\nu \bar{H}^{\sigma\tau}\bar{H}^{\rho u}\bar{H}^{rs}K_{\sigma\tau r}K_{\rho u s} \\
&+ (\nu+2)\sum_1^\nu \bar{H}^{rs}\bar{H}^{\sigma\tau}K_{\sigma\tau r}\sum_{\alpha=\nu+1}^{n} H_{s\alpha}\lambda^\alpha \\
&+ \frac{1}{4}(\nu+2)^2\sum_1^\nu \bar{H}^{rs}\sum_{\alpha,\beta=\nu+1}^{n} H_{r\alpha}H_{s\beta}\lambda^\alpha\lambda^\beta,
\end{aligned}
$$

$$
(34)\quad \begin{aligned}
\mathfrak{B} = {}& \sum_1^\nu \bar{H}^{\sigma\tau}\bar{H}^{\rho u}H_{\sigma\tau\rho u} + 4\sum_1^\nu \bar{H}^{\sigma\tau}\bar{H}^{\rho u}K_{\rho u\sigma}\sum_{\alpha=\nu+1}^{n} H_{\tau\alpha}\lambda^\alpha \\
&+ 2(\nu+2)\sum_1^\nu \bar{H}^{rs}\sum_{\alpha,\beta=\nu+1}^{n} H_{r\alpha}H_{s\beta}\lambda^\alpha\lambda^\beta + 2(\nu+2)\sum_1^\nu \bar{H}^{\sigma\tau}\sum_{\alpha=\nu+1}^{n} K_{\sigma\tau\alpha}\lambda^\alpha \\
&+ \frac{1}{2}\nu(\nu+2)\sum_{\alpha,\beta=\nu+1}^{n} H_{\alpha\beta}\lambda^\alpha\lambda^\beta,
\end{aligned}
$$

remembering that $H_{\sigma\tau}$, $K_{\sigma\tau\rho}$, $H_{\sigma\tau\rho u}$ are symmetrical in the subscripts and that

$$
(35)\quad \sum_1^\nu \bar{H}^{\sigma\tau}H_{\sigma\rho} = \delta_\rho^\tau .
$$

It remains for us to compute $\mathfrak{C}$. For this purpose we find it convenient to remark that

$$
(36)\quad \left\{\begin{aligned}
&\bar{H}^{\sigma u,\tau\rho} = -\bar{H}^{u\sigma,\tau\rho} = -\bar{H}^{\sigma u,\rho\tau} \\
&\bar{H}^{\sigma u,\tau\rho}H_{\tau r} = \bar{H}^{u\rho}\delta_r^\sigma - \bar{H}^{\sigma\rho}\delta_r^u , \\
&\bar{H}^{\sigma u,\tau\rho}H_{\sigma r} = \bar{H}^{u\rho}\delta_r^\tau - \bar{H}^{u\tau}\delta_r^\rho , \\
&\sum_1^\nu \bar{H}^{\sigma u,\tau\rho}(H_{\sigma\tau}H_{u\rho} - H_{u\tau}H_{\sigma\rho}) = 2\nu(\nu-1), \\
&\sum_1^\nu \bar{H}^{\sigma u,\tau\rho}(H_{\sigma r}H_{us} - H_{ur}H_{\sigma s}) = 2(\delta_r^\tau\delta_s^\rho - \delta_s^\tau\delta_r^\rho), \qquad (r, s, \rho, \tau = 1, \cdots, \nu).
\end{aligned}\right.
$$

The first two relations are consequences of the definition of $\bar{H}^{\sigma u,\tau\rho}$, the next two follow from the expansions of the determinant involved in $\bar{H}^{\rho u}$, and the last two sets follow from the repeated utilization of Laplace expansions of the determinant $\bar{H}$ or from the preceding two equations. Some calculations suffice then to demonstrate that

$$
\begin{aligned}
\mathfrak{C} = \sum_{1}^{\nu} \bar{H}^{rs}\bar{H}^{\sigma u,\tau\rho}(K_{\sigma\tau r}K_{u\rho s} - K_{u\tau r}K_{\sigma\rho s}) \\
+ 2(\nu - 1)\sum_{1}^{\nu}\bar{H}^{rs}\bar{H}^{\sigma\tau}K_{\sigma\tau r}\sum_{\alpha=\nu+1}^{n} H_{s\alpha}\lambda^{\alpha} \\
+ \frac{1}{2}(\nu + 2)(\nu - 1)\sum_{1}^{\nu}\bar{H}^{rs}\sum_{\alpha,\beta=\nu+1}^{n} H_{r\alpha}H_{s\beta}\lambda^{\alpha}\lambda^{\beta}.
\end{aligned} \tag{37}
$$

Substituting (33), (34) and (37) into (32) and reducing, we arrive at the equation to the quadric of Čech of V_ν at O, namely,

$$
\begin{aligned}
X^{n+1} - \frac{1}{2}\sum_{1}^{\nu} H_{\sigma\tau}X^{\sigma}X^{\tau} - \left\{\frac{2}{\nu+2}\sum_{1}^{\nu}\bar{H}^{\rho u}K_{\rho u\sigma}X^{\sigma} + \sum_{\sigma=1}^{\nu}\sum_{\alpha=\nu+1}^{n} H_{\sigma\alpha}\lambda^{\alpha}X^{\sigma}\right\}X^{n+1} \\
+ \left\{\frac{2(\nu+3)}{\nu(\nu+2)^2}\sum_{1}^{\nu}\bar{H}^{rs}\bar{H}^{\rho u}\bar{H}^{\sigma\tau}K_{\sigma\tau r}K_{\rho u s} - \frac{1}{\nu(\nu+2)}\sum_{1}^{\nu}\bar{H}^{\sigma\tau}\bar{L}_{\sigma\tau}\right. \\
+ \frac{4}{\nu(\nu+2)}\sum_{1}^{\nu}\bar{H}^{\sigma\tau}\bar{H}^{\rho u}K_{\rho u\sigma}\sum_{\alpha=\nu+1}^{n} H_{\tau\alpha}\lambda^{\alpha} - \frac{2}{\nu}\sum_{1}^{\nu}\bar{H}^{\sigma\tau}\sum_{\alpha=\nu+1}^{n} K_{\sigma\tau\alpha}\lambda^{\alpha} \\
\left. - \frac{1}{2}\sum_{\alpha,\beta=\nu+1}^{n} H_{\alpha\beta}\lambda^{\alpha}\lambda^{\beta}\right\}(X^{n+1})^2 = 0
\end{aligned} \tag{38}
$$

where $\bar{L}_{\sigma\tau}(\sigma, \tau = 1, \cdots, \nu)$ are given by (4). Elimination of $\lambda^{\nu+1}, \cdots, \lambda^{n}$ between (26) and (38) immediately gives *the equation of the Moutard-Čech hyperquadric belonging to the space* (25):

$$
\begin{aligned}
X^{n+1} = \frac{1}{2}\sum_{\sigma,\tau=1}^{n} H_{\sigma\tau}\cdot X^{\sigma}X^{\tau} \\
+ \sum_{\rho=1}^{n}\left\{\frac{2}{\nu}\sum \bar{H}^{\sigma\tau}K_{\sigma\tau\rho} - \frac{4}{\nu(\nu+2)}\sum_{\sigma,\tau,u,v=1}^{\nu} H_{\sigma\rho}\bar{H}^{\sigma\tau}\bar{H}^{uv}K_{uv\tau}\right\}X^{\rho}X^{n+1} \\
- \left\{\frac{2(\nu+3)}{\nu(\nu+2)}\sum_{1}^{\nu}\bar{H}^{\sigma\tau}\bar{H}^{\rho u}\bar{H}^{rs}K_{\sigma\tau r}K_{\rho u s} - \frac{1}{\nu(\nu+2)}\sum_{1}^{\nu}\bar{H}^{\sigma\tau}\bar{L}_{\sigma\tau}\right\}(X^{n+1})^2.
\end{aligned} \tag{39}
$$

Several interesting theorems can be easily established. For example, take any point $P(\xi^1, \cdots, \xi^\nu, 0, \cdots, 0)$ in the given space $[\nu]$; the polar hyperplane (in $[\nu]$) of P with respect to the corresponding Moutard-Čech hyperquadric is evidently given by the equation

$$
X^{n+1} = \sum_{\sigma,\tau=1}^{\nu} H_{\sigma\tau}\xi^{\sigma}X^{\tau} + \frac{2}{\nu+2}\sum_{1}^{\nu}\bar{H}^{\sigma\tau}K_{\sigma\tau\rho}\xi^{\rho}X^{n+1}. \tag{40}
$$

This hyperplane can, however, be obtained by the following method: Consider the quadric Q given by (20); the cubic hypercone Γ^3_{n-1} of equation (21) contains the space $[\nu]$ when and only when

$$\sum_1^\nu (\tfrac{1}{3}K_{\sigma\tau\rho} - \tfrac{1}{2}a_\rho H_{\sigma\tau})\xi^\sigma\xi^\tau\xi^\rho = 0,$$

that is,

$$K_{\sigma\tau\rho} - \tfrac{1}{2}a_\rho H_{\sigma\tau} - \tfrac{1}{2}a_\sigma H_{\tau\rho} - \tfrac{1}{2}a_\tau H_{\rho\sigma} = 0, \qquad (\sigma, \tau, \rho = 1, \cdots, \nu). \tag{41}$$

Multiplying (41) by $\bar{H}^{\sigma\tau}$ and summing for $\sigma, \tau = 1, \cdots, \nu$, we obtain

$$a_\rho = \frac{2}{\nu + 2}\sum_{\sigma,\tau=1}^{\nu} \bar{H}^{\sigma\tau} K_{\sigma\tau\rho}, \qquad (\rho = 1, \cdots, \nu). \tag{42}$$

The hyperquadric Q in this case becomes

$$\begin{aligned} X^{n+1} = \frac{1}{2}\sum_1^\nu H_{\sigma\tau}X^\sigma X^\tau + \frac{2}{\nu+2}\sum_1^\nu \bar{H}^{\sigma\tau}K_{\sigma\tau\rho}X^\rho X^{n+1} \\ + \sum_{\alpha=\nu+1}^{n} a_\alpha X^\alpha X^{n+1} + a(X^{n+1})^2, \end{aligned} \tag{43}$$

$a_{\nu+1}, \cdots, a_n$, a being arbitrary constants. It is obvious that the polar hyperplane of the point $P(\xi^1, \cdots, \xi^\nu, 0, \cdots, 0)$ is precisely (40). Hence we have the

THEOREM. *Suppose that a space $[\nu]$ through O be contained in the tangent hyperplane of a hypersurface V_n at O and that a hyperuqadric Q has at O a contact of the second order with V_n. If the cubic hypercone Γ^3_{n-1} formed by the tangents drawn from O to the intersection of V_n and Q contains the space $[\nu]$, then the polar hyperplane of any point in this $[\nu]$ with respect to Q must coincide with that of the same point with respect to the Moutard-Čech hyperquadric belonging to $[\nu]$.*

In particular when $\nu = 1$ we obtain a theorem due to Fubini.[9]

Let us now consider the hyperquadric (20) such that the corresponding Γ^3_{n-1} should pass through the space $[\nu]$ doubly. We can show that such a hyperquadric must belong to a pencil. In fact, for the hyperquadric Q in question the coefficients $a_\rho(\rho = 1, \cdots, \nu)$ are, as before, given by (42), and

$$\frac{\partial F_3}{\partial X^\rho} \equiv (K_{\sigma\tau\rho} - \tfrac{1}{2}a_\rho H_{\sigma\tau} - a_\sigma H_{\tau\rho})X^\sigma X^\tau = 0, \quad (\rho = 1, \cdots, n),$$

must be satisfied identically for any $\xi^1, \cdots, \xi^\nu$, and $\xi^\alpha = 0$, $(\alpha = \nu + 1, \cdots, n)$. Therefore

$$\sum_{\sigma,\tau=1}^{\nu} (K_{\sigma\tau\rho} - \tfrac{1}{2}a_\rho H_{\sigma\tau} - a_\sigma H_{\tau\rho})X^\sigma X^\tau \equiv 0, \quad (\rho = 1, \cdots, n),$$

or

$$2K_{\sigma\tau\rho} - a_\rho H_{\sigma\rho} - a_\sigma H_{\tau\rho} - a_\tau H_{\sigma\rho} = 0 \quad (\sigma, \tau = 1, \cdots, \nu; \rho = 1, \cdots, n), \tag{44}$$

---

[9] Cf. *G. P. D.*, pp. 617–618.

whence

$$(45)\quad a_\rho = \frac{2}{\nu}\sum_{\sigma,\tau=1}^{\nu} \bar{H}^{\sigma\tau}K_{\sigma\tau\rho} - \frac{4}{\nu(\nu+2)}\sum_{\sigma,\tau,u,v=1}^{\nu} H_{\rho\sigma}\bar{H}^{\sigma\tau}\bar{H}^{uv}K_{uv\tau} \quad (\rho = 1, \cdots, n).$$

Hence we obtain a pencil of hyperquadrics

$$(46)\quad \begin{aligned} X^{n+1} &= \frac{1}{2}\sum_{\sigma,\tau=1}^{n} H_{\sigma\tau}X^\sigma X^\tau \\ &+ \sum_{\rho=1}^{n}\left\{\frac{2}{\nu}\sum_{\sigma,\tau=1}^{\nu}\bar{H}^{\sigma\tau}K_{\sigma\tau\rho} - \frac{4}{\nu(\nu+2)}\sum_{\sigma,\tau,u,v=1}^{\nu} H_{\rho\sigma}\bar{H}^{\sigma\tau}\bar{H}^{uv}K_{uv\tau}\right\}X^\rho X^{n+1} \\ &+ a(X^{n+1})^2, \end{aligned}$$

a being a parameter. We shall call (46) *the pencil of Moutard-Čech* belonging to the given space $[\nu]$, because each of these pencils contains a Moutard-Čech hyperquadric (39). Thus we have proved the following

THEOREM. *If the cubic hypercone Γ^3_{n-1} passes through the space $[\nu]$ doubly, then the hyperquadric Q must necessarily be in a pencil of Moutard-Čech belonging to the space $[\nu]$.*

That the converse of this theorem is not necessarily true may be seen by considering a V_n in which $H_{\sigma\tau} = \delta_{\sigma\tau}$ and $K_{\rho\sigma\tau} \neq 0$ for ρ, σ, τ all different. Then even if a_ρ is defined by (45), the right member of (44) becomes $K_{\rho\sigma\tau} \neq 0$ for ρ, σ, τ all different, and thus (44) cannot hold.

This generalizes a theorem for the case $\nu = 1$.[10]

The hyperquadric of the pencil (46) possesses another definition, as we will show below.

From (27) it is easily seen that all of sections V_ν produced by spaces $[\nu + 1]$ through the space $[\nu]$ have the asymptotic hypercone at O in common, namely,

$$(47)\quad \sum_{\sigma,\tau=1}^{\nu} H_{\sigma\tau}X^\sigma X^\tau = 0, \qquad X^{\nu+1} = 0, \cdots, X^{n+1} = 0,$$

which may be obtained as the intersection of the space $[\nu]$ with the asymptotic hypercone (22).

In the tangent hyperplane $X^{n+1} = 0$ of V_n at O a space $[n - \nu - 1]$ is taken such that it is skew with the space $[\nu]$. The projection of the hypercone (47) from this $[n - \nu - 1]$ is *apolar* to the cubic hypercone Γ^3_{n-1} if and only if

$$\sum_{\sigma,\tau=1}^{\nu} \bar{H}^{\sigma\tau}(2K_{\sigma\tau\rho} - a_\rho H_{\sigma\tau} - a_\sigma H_{\rho\tau} - a_\tau H_{\rho\sigma}) = 0, \qquad (\rho = 1, \cdots, n),$$

whence we obtain (45) and consequently the pencil of hyperquadrics (46). Hence we have the

THEOREM. *If the cubic hypercone Γ^3_{n-1} be apolar to the quadratic cone obtained by projecting the common asymptotic hypercone of various V_ν's from a space*

[10] Cf. my paper: *Plane sections through an ordinary point of a hypersurface*, to be published in Revista, Tucumán.

$[n - \nu - 1]$ *contained in the tangent hyperplane of* V_n *at* O, *but skew with the given* $[\nu]$, *then the corresponding hyperquadric* Q *must belong to the pencil of Moutard-Čech, and conversely.*

We have shown the truth of this theorem in the case $\nu = 2$.[11]

Among various sections of a hyperquadric Q of the equation (20) produced by spaces (26) of $\nu + 1$ dimensions through the given space $[\nu]$ it may happen that the section just coincides with the quadric of Čech of the section V_ν. In virtue of (20) and (26) it is easily shown that the section of Q is

$$(48)\qquad \begin{aligned} 2X^{n+1} = {} & \sum_{\sigma,\tau=1}^{\nu} H_{\sigma\tau} X^\sigma X^\tau + 2\sum_{\sigma=1}^{\nu}\left\{a_\sigma + \sum_{\alpha=\nu+1}^{n} H_{\sigma\alpha}\lambda^\alpha\right\} X^\sigma X^{n+1} \\ & + 2\left(a + \sum_{\alpha=\nu+1}^{n} a_\alpha \lambda^\alpha\right)(X^{n+1})^2. \end{aligned}$$

In order that the latter should represent the quadric of Čech (38), the necessary and sufficient conditions are

$$(49)\qquad a_\sigma = \frac{2}{\nu+2}\sum_{\rho,u=1}^{\nu} \bar{H}^{\rho u} K_{\rho u\sigma}, \qquad (\sigma = 1, \cdots, \nu),$$

$$(50)\qquad \begin{aligned} a + \sum_{\alpha=\nu+1}^{n} a_\alpha \lambda^\alpha = {} & \frac{1}{2}\sum_{\alpha,\beta=\nu+1}^{n} H_{\alpha\beta}\lambda^\alpha\lambda^\beta + \frac{2}{\nu}\sum_{1}^{\nu} \bar{H}^{\sigma\tau} \sum_{\alpha=\nu+1}^{n} K_{\sigma\tau\alpha}\lambda^\alpha \\ & + \frac{1}{\nu(\nu+2)} \sum^{\nu} \bar{H}^{\sigma\tau} \bar{L}_{\sigma\tau} \\ & - \frac{4}{\nu(\nu+2)} \sum_{1}^{\nu} \bar{H}^{\sigma\tau}\bar{H}^{\rho u} K_{\rho u\sigma} \sum_{\alpha=\nu+1}^{n} H_{\tau\alpha}\lambda^\alpha \\ & - \frac{2(\nu+3)}{\nu(\nu+2)^2} \sum_{1}^{\nu} \bar{H}^{\tau s}\bar{H}^{\sigma\tau}\bar{H}^{\rho u} K_{\sigma\tau r} K_{\rho u s}. \end{aligned}$$

I have not been able to find the *geometrical meaning* of (49) and (50). We remark that if the hypercone Γ^3_{n-1} passes through $[\nu]$ then (42), which is identical with (49), holds. However, since (49), or (42), does not imply (41)—as can be seen from the counterexample following the next to the last theorem above—(49) does not imply that Γ^3_{n-1} passes through $[\nu]$. We also remark that in the case $\nu = 1$ the equation (50) shows that *the plane* (26) *should osculate a curve contained in the intersection of* V_n *and* Q.[12]

It is of some interest to investigate all the quadrics of Čech of sections V_ν which lie on a fixed hyperquadric Q having at O a contact of the second order with V_n. Observing that a space $[\nu]$ through O and in the tangent hyperplan of V_n at O contains in general $\nu(n - \nu)$ parameters and that the conditions (49) and (50) are $\nu + 1$ in number, we may conclude that *on* Q *there are* $\infty^{(\nu+1)(n-\nu-1)}$ *quadrics of Čech of* ν *dimensions.*

---

[11] Cf. my paper, loc. cit.[10].

[12] Cf. my paper, loc. cit.[10].

For the sake of completeness we shall find the equation of the Moutard-Čech hyperquadric belonging to a general space $[\nu]$ through O and in the tangent hyperplane of V_n at O.

For this purpose it is convenient to express the equations to the space $[\nu]$ in the form

$$X^\alpha = \sum_{\sigma=1}^{\nu} d_\sigma^\alpha X^\sigma \qquad (\alpha = \nu + 1, \cdots, n), \qquad X^{n+1} = 0 \tag{51}$$

or

$$\bar{X}^k = 0 \qquad (k = \nu + 1, \cdots, n), \qquad X^{n+1} = 0 \tag{52}$$

if we put

$$\bar{X}^k = X^k - \sum_{\sigma=1}^{\nu} d_\sigma^k X^\sigma \qquad (k = \nu + 1, \cdots, n). \tag{53}$$

Introduce

$$\bar{X}^j = X^j \qquad (j = 1, \cdots, \nu), \qquad \bar{X}^{m+1} = X^{n+1}, \tag{54}$$

so that

$$X^k = \bar{X}^k + \sum_{\sigma=1}^{\nu} d_\sigma^k X^\sigma \qquad (k = \nu + 1, \cdots, n). \tag{55}$$

Substituting (54), (55) into (1) and rearranging, we obtain

$$\begin{aligned} \bar{X}^{n+1} &= \frac{1}{2}\sum_1^n \bar{H}_{\sigma\tau}\bar{X}^\sigma \bar{X}^\tau + \frac{1}{3}\sum_1^n \bar{K}_{\sigma\tau\rho}\bar{X}^\sigma\bar{X}^\tau\bar{X}^\rho \\ &\qquad + \frac{1}{12}\sum \bar{H}_{\sigma\tau\rho u}\bar{X}^\sigma\bar{X}^\tau\bar{X}^\rho\bar{X}^u + \cdots, \end{aligned} \tag{56}$$

where

$$\left\{\begin{aligned} \bar{H}_{\sigma\tau} &= H_{\sigma\tau} + \sum_{\alpha=\nu+1}^{n} H_{\sigma\alpha} d_\tau^\alpha + \sum_{\alpha=\nu+1}^{n} H_{\tau\alpha} d_\sigma^\alpha + \sum_{\alpha,\beta=\nu+1}^{n} H_{\alpha\beta} d_\sigma^\alpha d_\tau^\beta, \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\sigma, \tau = 1, \cdots, \nu), \\ \bar{H}_{\sigma\alpha} &= H_{\sigma\alpha} + \sum_{\beta=\nu+1}^{n} H_{\alpha\beta} d_\sigma^\beta, \qquad (\sigma = 1, \cdots, \nu; \alpha = \nu + 1, \cdots, n), \\ \bar{H}_{\alpha\beta} &= H_{\alpha\beta}, \qquad\qquad\qquad\qquad (\alpha, \beta = \nu + 1, \cdots, n); \end{aligned}\right. \tag{57}$$

and similarly one can obtain explicit expressions for $\bar{K}_{\rho\sigma\tau}$ and $\bar{H}_{\rho\sigma\tau u}$ in terms of $K_{\rho\sigma\tau}$, $H_{\rho\sigma\tau u}$, and d_ρ^α.

By virtue of (39) the equation of the Moutard-Čech hyperquadric belonging to (51) is of the form

$$\begin{aligned} \bar{X}^{n+1} &= \frac{1}{2}\sum_1^n \bar{H}_{\sigma\tau}\bar{X}^\sigma\bar{X}^\tau \\ &\quad + \sum_{\rho=1}^{n}\left\{\frac{2}{\nu}\sum_1^\nu \bar{\bar{H}}^{\sigma\tau}\bar{K}_{\sigma\tau\rho} - \frac{4}{\nu(\nu+2)}\sum_1^\nu \bar{H}_{\sigma\rho}\bar{\bar{H}}^{\sigma\tau}\bar{\bar{H}}^{uv}\bar{K}_{uv\tau}\right\}\bar{X}^\rho\bar{X}^{n+1} \\ &\quad - \left\{\frac{2(\nu+3)}{\nu(\nu+2)^2}\sum_1^\nu \bar{\bar{H}}^{\tau s}\bar{\bar{H}}^{\sigma\tau}\bar{\bar{H}}^{\rho u}\bar{K}_{\sigma\tau\tau}\bar{K}_{\rho u s} - \frac{1}{\nu(\nu+2)}\sum_1^\nu \bar{\bar{H}}^{\sigma\tau}\bar{L}_{\sigma\tau}\right\}(\bar{X}^{n+1})^2 \end{aligned} \tag{58}$$

where $\bar{\bar{H}}^{\sigma\tau}$ is related to $\bar{H}_{\sigma\tau}$ as $\bar{H}_{\sigma\tau}$ was to $H_{\sigma\tau}$. On account of (53), (54), (57) the equation (58) may be written as

$$
\begin{aligned}
X^{n+1} &= \frac{1}{2}\sum_1^n H_{\sigma\tau}X^\sigma X^\tau + \frac{2}{\nu}\sum_{\rho=1}^n\Big\{\sum_1^\nu \bar{\bar{H}}^{\sigma\tau}\bar{K}_{\sigma\tau\rho} - \frac{2}{\nu+2}\sum_1^\nu \bar{H}_{\sigma\rho}\bar{\bar{H}}^{\sigma\tau}\bar{\bar{H}}^{uv}\bar{K}_{uv\tau} \\
&\quad - \sum_{\alpha=\nu+1}^n\Big(\sum_1^\nu \bar{\bar{H}}^{\sigma\tau}\bar{K}_{\sigma\tau\alpha} - \frac{2}{\nu+2}\sum_1^\nu \bar{H}_{\sigma\alpha}\bar{\bar{H}}^{\sigma\tau}\bar{\bar{H}}^{uv}\bar{K}_{uv\tau}\Big)d_\rho^\alpha\Big\}X^\rho X^{n+1} \qquad (59) \\
&\quad + \frac{1}{\nu(\nu+2)}\Big\{\sum_1^\nu \bar{\bar{H}}^{\sigma\tau}\bar{L}_{\sigma\tau} - 2(\nu+3)\sum_1^\nu \bar{\bar{H}}^{rs}\bar{\bar{H}}^{\sigma\tau}\bar{\bar{H}}^{\rho u}\bar{K}_{\sigma\tau r}\bar{K}_{\rho us}\Big\}(X^{n+1})^2.
\end{aligned}
$$

Hence we have the

THEOREM. *The Moutard-Čech hyperquadrics belonging to two different spaces through O and in the tangent hyperplane of V_n at O intersect in the asymptotic hypercone of V_n at O and a hyperquadric of $n - 1$ dimensions.*

We shall consider now the intersection of the quadric of Čech of V_n with the Moutard-Čech hyperquadric (59) or what amounts to the same thing, (58). Since (19) remains unaltered when H, K, X are replaced by $\bar{H}$, $\bar{K}$, $\bar{X}$ respectively, the intersection in question consists of the asymptotic hypercone and a hyperquadric Q_{n-1}. The hyperplane on which Q_{n-1} lies passes through the space (52) if and only if

$$
(60) \qquad \frac{1}{\nu+2}\sum_1^\nu \bar{\bar{H}}^{\rho u}\bar{K}_{\rho u\sigma} - \frac{1}{n+2}\sum_1^n \bar{H}^{\rho u}\bar{K}_{\rho u\sigma} = 0, \qquad (\sigma = 1, \cdots, \nu).
$$

But the Darboux tangents of V_n at 0 constitute a cubic hypercone with the equations

$$
(61) \qquad \sum_1^n \mathfrak{K}_{\rho u\sigma}\bar{X}^\rho\bar{X}^u\bar{X}^\sigma = 0, \qquad \bar{X}^{n+1} = 0,
$$

where we have placed[13]

$$
(62) \qquad
\begin{aligned}
\mathfrak{K}_{\rho u\sigma} = \bar{K}_{\rho u\sigma} - \frac{1}{n+2}\Big(&\bar{H}_{u\sigma}\sum_1^n \bar{H}^{rs}\bar{K}_{rs\rho} \\
&+ \bar{H}_{\rho\sigma}\sum_1^n \bar{H}^{rs}\bar{K}_{rsu} + \bar{H}_{\rho u}\sum_1^n \bar{H}^{rs}\bar{K}_{rs\sigma}\Big).
\end{aligned}
$$

In order that this hypercone should contain the space (52) it is necessary and sufficient that

$$
(63) \qquad \mathfrak{K}_{\rho u\sigma} = 0 \qquad (\rho, u, \sigma = 1, \cdots, \nu)
$$

Multiplying (63) by $\bar{\bar{H}}^{\rho u}$ and summing up with regard to ρ, $u = 1, \cdots, \nu$, we are led to (60). Hence there follows the

---

[13] Cf. Kanitani, loc. cit., p. 34.

THEOREM. *At a generic point O of a hypersurface V_n the quadric of Čech (or any quadric in the Darboux pencil) and the Moutard-Čech hyperquadric (or any hyperquadric in the pencil of Moutard-Čech) belonging to a tangent space $[\nu]$ intersect in the asymptotic hypercone of V_n and another hyperquadric of $n - 1$ dimensions. The hyperplane through the latter hyperquadric contains the space $[\nu]$ when and only when the space $[\nu]$ belongs to the cubic hypercone of Darboux tangents at O.*

Thus is generalized the theorem of Bompiani[14] concerning Moutard quadrics of a surface in ordinary space.

NATIONAL UNIVERSITY OF CHEKIANG,
KWEICHOW, CHINA.

[14] E. Bompiani, *Contributi alla geometria proiettiva-differenziale di una superficie*, Bollettino della Unione Matematical Italiana, 3 (1924), p. 97.

Annals of Mathematics
Vol. 44, No. 1, January, 1943

THE CLASS-RING IN MULTIPLICATIVE SYSTEMS

By A. R. Richardson

(Received October 15, 1942)

Notation. Σ denotes a system closed to a binary operation; $a, b, c, \cdots$, are elements of E-subsets $A, B, C, \cdots$ of Σ; $u, v, w, \cdots$, are elements of Σ. The term E-set implies that the classes $C_w = w \cdot A$ are disjoint, the relation between the elements in C_w being an equivalence relation R_A. Hence if w be any element in Σ, a_1, a_2 any elements in A then there exists an element a such that $w = w \cdot a$; $w = (w \cdot a_1) \cdot a_2$; $(w \cdot a_1) \cdot a_2 = w \cdot a$.

These classes will be regarded as the elements of a multiplicative system T_A in which multiplication will be denoted by $\times$. In group theory $\times$ may be derived from $\cdot$, e.g. $a \times b$ may denote any one of $a \cdot b$, $b \cdot a$, $a \cdot b \cdot a^{-1} \cdot b^{-1}$, $b^{-1} \cdot a \cdot b$ etc. On the other hand $\times$ may be defined without reference to $\cdot$ as in the theory of rings of systems closed to two operations, e.g. $ab - ba$, $ab + ba$.[1]

In order to establish an automorphism between the structure (lattice) of equivalence relations R_A, in which A ranges over all E-subsets of Σ, and those quotients of E-sets having the same R_A, we shall assume:

Cancellation Law I. *If for any element u we have $u \cdot a = u \cdot b$ then there exists a d in $D = A \cap B$ such that $u \cdot a = u \cdot b = u \cdot d$, i.e. we assume that D is non-void.*

Associative Laws. $w \cdot (a \cdot b) = (w \cdot a) \cdot \bar{b}$; $(w \cdot a) \cdot b = w \cdot (a \cdot \bar{b})$ *where $\bar{b}$ is in B. Hence $A \cdot (B \cdot C) = (A \cdot B) \cdot C$.*

The product $w \cdot a_i \times v \cdot a_j$ is in Σ and so falls into one and only one class $C_u = u \cdot A$. In ordinary algebra several classes arise, viz.:

(a) All elements $w \cdot A \times v \cdot A$ are in the same class $(w \times v) \cdot A$. This holds in the theory of normal *co*-set expansions in groups and semi-groups.

(b) All elements $w \cdot A \times v \cdot A$ are in the same class $u \cdot A$ where $u \neq w \times v$. This case cannot occur here for since A is an E-set, right units a_w, a_v exist such that $w = w \cdot a_w$ and $v = v \cdot a_v$. Hence $w \times v$ is in $w \cdot A \times v \cdot A$ which product must therefore be in $(w \times v) \cdot A$.

(c) The elements $w \cdot A \times v \cdot A$ fall into different classes, e.g. the classes of conjugate elements in group theory.

We shall denote unions and cross-cuts with respect to $\cdot$ and $\times$ by $A \overset{\cdot}{\cup} B$, $A \overset{\times}{\cup} B$, $A \overset{\cdot}{\cap} B$, $A \overset{\times}{\cap} B$ respectively.

1. $C_w \times C_v \leqq C_{u \times v}$.

This is equivalent to

$$(w \cdot A) \times (v \cdot A) \leqq (w \times v) \cdot A \tag{1}$$

[1] A. R. Richardson, *Congruences in Multiplicative Systems*. To appear in Proc. London Math. Soc.

i.e. given a_i, a_j any elements of A; w, v, any elements of Σ then $a \in A$ exists such that

$$(w \cdot a_i) \times (v \cdot a_j) = (w \times v) \cdot a. \tag{2}$$

Since A is an E-set $a_w \in A$ exists such that $w \cdot a_w = w$. Hence there are the following special instances of the distributive law (2):

$$w \times (v \cdot a_j) = (w \times v) \cdot a \tag{3}$$

$$(w \cdot a_i) \times v = (w \times v) \cdot a.[2] \tag{4}$$

If $(w \cdot A) \times (v \cdot A) = (w \times v) \cdot A$ there are other special instances:

$$w \times (a_i \cdot a_j) = (w \times a_i) \cdot a_j,$$

$$(w \times a_i) \cdot a_j = (w \cdot a_k) \times (a_i \cdot a_m) = (w \cdot a_k) \times a_n,$$

$$(w \cdot a_i) \times a_j = (w \cdot a_i) \times (a_j \cdot a_l) = (w \times a_j) \cdot a.$$

THEOREM 1. *If the classes mod R_A and mod R_B both satisfy condition* (2) *then so do those mod $R_{A \dot{\cup} B}$ and mod $R_{A \dot{\cap} B}$.*

Let $D = A \dot{\cap} B$, supposedly non-void, and let d_i, d_j be any elements in D. Then $(w \cdot d_i) \times (v \cdot d_j) = (w \cdot a_i) \times (v \cdot a_j) = (w \times v) \cdot a$. Similarly $(w \cdot d_i) \times (v \cdot d_j) = (w \times v) \cdot b$. Hence, by cancellation law I, there exists d in D such that $(w \cdot d_i) \times (v \cdot d_j) = (w \times v) \cdot d$, i.e. (2) holds for $D = A \dot{\cap} B$. Next, owing to the associative laws, the elements of $M = A \dot{\cup} B$ are of one of the forms a, b, $a \cdot b$, $b \cdot a$. Take these in turn:

$$\begin{aligned}(w \cdot a_i) \times (v \cdot b_j) &= \{(w \cdot a_i) \times v\} \cdot b = \{(w \times v) \cdot a\} \cdot b = (w \times v) \cdot (a \cdot \bar{b}) \\ &= (w \times v) \cdot m.\end{aligned}$$

$$\begin{aligned}\{w \cdot (a_1 \cdot b_1)\} \times \{v \cdot (a_2 \cdot b_2)\} &= \{(w \cdot a_1) \cdot b_3\}\{(v \cdot a_2) \cdot b_4\} = \{(w \cdot a_1) \times (v \cdot a_2)\} \cdot b \\ &= \{(w \times v) \cdot a\} \cdot b = (w \times v) \cdot (a \cdot b_5) = (w \times v) \cdot m.\end{aligned}$$

$$\begin{aligned}\{w \cdot (a_1 \cdot b_1)\} \times (v \cdot (b_2 \cdot a_2)\} &= \{w \cdot (a_1 \cdot b_1)\} \times \{(v \cdot b_2) \cdot a_3\} \\ &= [\{w \cdot (a_1 \cdot b_1)\} \times (v \cdot b_2)] \cdot a = [\{(w \cdot a_1) \cdot b_3\} \times \{v \cdot b_2\}] \cdot a = [\{(w \cdot a_1) \times v\} \cdot b] \cdot a \\ &= [\{(w \times v) \cdot a_4\} \cdot b] \cdot a = (w \times v) \cdot m.\end{aligned}$$

Hence (2) holds for $M = A \dot{\cup} B$.

COROLLARY 1. *We have also proved that $A \cdot B$, defined as the set of all elements of Σ of the form $a \cdot b$, also satisfies* (2).

COROLLARY 2. *If A is a normal E-set of Σ with respect to $\cdot$, i.e. if $v \cdot A = A \cdot v$ for all v in Σ, then* (2) *is satisfied if $\times$ is taken to be $\cdot$.*

If (2) is satisfied it is customary to define $C_w \times C_v$ to be equal to $C_{w \times v}$, for every

[2] These laws resemble those of D. C. Murdock and O. Ore, *On generalized rings*. Am. Journ. of Math. vol. LXIII No. 1. 1941.

element in $C_w \times C_v$ is in $C_{w \times v}$. In such cases the correspondence $w \to w \cdot A$ between Σ and T_A is a homomorphism.

Although in many instances we require only that $(w \cdot A) \times (v \cdot A) \leqq (w \times v) \cdot A$ in others it is necessary that $(w \cdot A) \times (v \cdot A) = (w \times v) \cdot A$. We therefore examine the condition $(w \times v) \cdot A \leqq (w \cdot A) \times (v \cdot A)$. For this to hold it is necessary that given w, v, a there shall exist a_i and a_j such that

$$(w \times v) \cdot a = (w \cdot a_i) \times (v \cdot a_j). \tag{5}$$

THEOREM 2. *If the classes mod R_A and mod R_B satisfy* (5) *then so do those mod $R_{A \dot{\cup} B}$.*

The elements of $A \dot{\cup} B$ are of one of the forms a, b, $a \cdot b$, $b \cdot a$. Consider

$$\begin{aligned}(w \times v) \cdot m &= (w \times v) \cdot (a \cdot b) = [(w \times v) \cdot a] \cdot b_2 = [(w \cdot a_i) \times (v \cdot a_j)] \cdot b_2 \\ &= \{(w \cdot a_i) \cdot b_i\} \times \{(v \cdot a_j) \cdot b_j\} = \{w \cdot (a_i \cdot \bar{b}_i)\} \times \{v \cdot (a_j \cdot \bar{b}_j)\} \\ &= (w \cdot m_i) \times (v \cdot m_j)\end{aligned}$$

i.e. (5) holds. Similarly it holds for the case $m = b \cdot a$ and therefore in all cases.

In order that (5) shall hold for $R_{A \dot{\cap} B}$ a new form of the cancellation law is necessary.

CANCELLATION LAW II. *If $(w \cdot a_i) \times (v \cdot a_j) = (w \cdot b_i) \times (v \cdot b_j)$ then there exist in $D = A \dot{\cap} B$ d_i and d_j such that the above products equal $(w \cdot d_i) \times (v \cdot d_j)$.*

THEOREM 3. *If the cancellation law II holds and if R_A, R_B satisfy condition* (5) *then so does $R_{A \dot{\cap} B}$.*

$(w \times v) \cdot d = (w \cdot a_i) \times (v \cdot a_j) = (w \cdot b_i) \times (v \cdot b_j)$ since d is in both A and B. Hence, from cancellation law II, d_i, d_j exist in D such that $(w \times v) \cdot d = (w \cdot d_i) \times (v \cdot d_j)$, i.e. (5) holds in $R_{A \dot{\cap} B}$.

THEOREM 4. *If Σ is homogeneous with respect to $\times$ and if cancellation law II and laws* (2) *and* (5) *hold for A and B then the cancellation law I also holds.*

Let z be any element in Σ. Since Σ is homogeneous, w and v exist such that $z = w \times v$. Let $z \cdot a = z \cdot b$ then a_i, a_j, b_i, b_j exist such that $z \cdot a = z \cdot b = (w \cdot a_i) \times (v \cdot a_j) = (w \cdot b_i) \times (v \cdot b_j)$ and by law II, d_i, d_j exist in $D = A \dot{\cap} B$ such that $z \cdot a = z \cdot b = (w \cdot d_i) \times (v \cdot d_j)$ and by law (2), there exists d in D such that these products are equal to $(w \times v) \cdot d = z \cdot d$, i.e. cancellation law I holds.

THEOREM 5. *If Σ is homogeneous with respect to Σ and if cancellation law I holds as well as laws* (2) *and* (5), *then the cancellation law II holds.*

Let $(w \cdot a_i) \times (v \cdot a_j) = (w \cdot b_i) \times (v \cdot b_j)$ then by (2) a and b exist such that $(w \times v) \cdot a = (w \times v) \cdot b$. Hence by cancellation law I, d exists in $D = A \dot{\cap} B$ such that $(w \times v) \cdot a = (w \times v) \cdot b = (w \times v) \cdot d$ and, since (5) also holds, this is equal to $(w \cdot d_i) \times (v \cdot d_j)$, i.e. cancellation law II holds.

In general the E-sets do not form a Dedekind Structure (modular lattice). If however the equation $a_1 = a_2 \cdot x$ is always solvable in an E-set, then

THEOREM 6. *The E-sets form a Dedekind structure with respect to $\cdot$, i.e. if $A < C$ then $C \dot{\cap} (A \dot{\cup} B) = A \dot{\cup} (B \dot{\cap} C)$.*

By the associate laws the elements of $A \dot{\cup} B$ are of one of the forms a, b, $a \cdot b$, $b \cdot a$.

Let $c = b \cdot a$. Then $\bar{a}$ exists in A and therefore in C such that $c \cdot \bar{a} = b$. Hence b is in C and therefore in $B \dot{\cap} C$, i.e. $e \in A \dot{\cup} (B \dot{\cap} C)$.

Let $c = a \cdot b$ then $\bar{c}$ exists in C such that $c = a \cdot \bar{c}$, i.e. $c = a \cdot \bar{c} = a \cdot b$. Hence d exists in $B \dot{\cap} C$ such that $c = a \cdot d$, i.e. $c \in A \dot{\cup} (B \dot{\cap} C)$.

COROLLARY. *The mutually permutable E-sets A, B, $\cdots$ for which $a_1 = a_2 \cdot x$, $b_1 = b_2 \cdot x$ have solutions, form a Dedekind structure.*

It is desirable to express the conditions that every element in a product class occurs equally often in the product but, apart from the very restrictive conditions such as those in group theory, the necessary qualifications are complicated and are not inserted here. Instead we shall assume the $C_i C_j = \Sigma c_{ijk} C_k$.

2. The characteristic equations in the class-ring

Let $X = \sum x_i C_i$, $i = 1, 2, \cdots, n$, the x's being indeterminates in a ring K not of characteristic two. Then

$$X \times Y = \sum x_i C_i \times \sum y_j C_j = \sum_{i,j} x_i y_j C_i C_j = (x_i)(\sum_j y_j c_{kjm})(C_n)$$

where (x_i) is a one-rowed matrix and C_n a one-columned matrix and $\Delta_Y = (\sum_j y_j c_{kjm})$ is an n-matrix in which k denotes the rows and m the columns. Hence

$$(x_i)(C_p)(y_j)(C_q) = (x_i)\Delta_Y(C_n).$$

Similarly

$$\begin{aligned} X \times Y &= (C_i)(\sum_j x_j c_{jmk})(y_n) \\ &= (y_i)(\sum x_j c_{jkm})(C_n) \\ &= (C_i)(\sum y_j c_{mjk})(x_n). \end{aligned}$$

Evidently $(X \times Y) \times Z = (x_i)\Delta_Y \Delta_Z (C_n)$ and $X \times (Y \times Z) = (x_i)\Delta_{Y \times Z}(C_n)$.

Hence the condition for the class-ring to be associative is $\Delta_{Y \times Z} = \Delta_Y \Delta_Z$. If this holds then the correspondence $Y \to \Delta_Y$ between the class-ring and the matrices in a homomorphism m in which the class multiplication x corresponds to matrix multiplication and addition to addition.

Let $X^{(n+1)} = X^{(n)} \times X$. Then $X^{(n+1)} = (x_i)\Delta_X^n(C_n)$. Hence if Δ_X satisfies the characteristic equation $Z^n + p_1 Z^{n-1} + \cdots + p_n = 0$, then X satisfies a characteristic equation

$$X^{(n+1)} + p_1 X^{(n)} + \cdots + p_n X = 0.$$

Evidently the p_i are homogeneous and of degree i in the x's and the elements of Δ_X are linear in the x's and integral in the c_{ijk}.

In group theory and in the theory of matrix representations of algebras the

regular representation is important. From the present point of view this corresponds to Δ_E where R_E is the unit equivalence relation in which each class consists of one and only one element of Σ.

THEOREM 7. *If every element in a class-product occurs equally often in its class then the characteristic polynomial of the general number in the class-ring has a linear factor* $Z - \sum x_i\rho_i$ *where* ρ_i *is the number of elements of* Σ *in* C_i.

Since every element occurs equally often in its class,

$$\rho_i\rho_j = \sum c_{ijk}\rho_k .$$

Multiply the columns of Δ_Z by $\rho_1, \rho_2, \cdots \rho_n$ respectively and add to the first. Then the γ^{th} row has in the first column the element

$$\sum_{j,m} x_j C_{\gamma jm}\rho_m - Z\rho_\gamma = \sum_j z_j\rho_\gamma\rho_j - Z\rho_\gamma = \rho_\gamma\left[\sum z_j\rho_j - Z\right]. \tag{6}$$

Hence $Z - \sum z_j\rho_j$ is a factor of the characteristic polynomial Δ_Z and it is linear in the indeterminates z_j.

In group theory $a\cdot b = b^{-1}\cdot a\cdot b$ and $a \times b = a\cdot b$. Every element in a class occurs equally often in a class-product and the characteristic equation of the general number in the class-ring splits into linear factors

$$\prod_i \left[Z - \sum_j z_j\rho_j X_j^{(i)}/x_0^{(i)}\right]$$

where the $X_j^{(i)}$ are the group characteristics. The group character $(1, 1, \cdots, 1)$ arises from (6).

In the general case the characteristic polynomial does not split into linear factors although it is known to do so when the class-ring is commutative.

THEOREM 8. *If* Σ *is an E-set with respect to* $\times$ *then* $\Delta_Z = [Z - \sum z_i\rho_i]^n$ *is a complete* n^{th} *power.*

For, Σ being an E-set, the elements which appear in the multiplicative table of Σ as multiples of C_i, $i = 1, 2, \cdots, n$, are in C_i. Hence $c_{ijk} = 0$, $k \neq i$ and $C_{iji} = \rho_j$, $k = i$. Hence $\Delta_Z = [Z - \sum z_i\rho_i]^n$.

THEOREM 9. *If* $R_A \subset R_B$ *then* $|\Delta_B|$ *is a divisor of* $|\Delta_A'|$ *where* Δ_A' *is what* Δ_A *becomes when certain of the a's are equal.*

Since $R_A \subset R_B$, $B_i = \sum C_s$ and therefore $B_iB_j = \sum b_{ijk}C_k = \sum C_nC_m = \sum c_{nmk}C_k$ i.e. $\sum b_{ijk} = \sum C_p = \sum c_{nmk}C_k$. Let $B_1 = C_1 + \cdots + C_{s_1}$; $B_2 = C_{s_1+1} + \cdots + C_{s_2}, \cdots$
then

$$B_iB_j = \sum (C_{s_{i-1}+1} + \cdots + C_{s_i})(C_{s_{j-1}+1} + \cdots + C_{s_j})$$
$$= \sum C_{s_{i-1}+j}C_{s_{j-1}+t} = \sum C_{s_{i-1}+\gamma,s_{j-1}+t_1p}C_p$$

i.e.

$$b_{ijp} = \sum_{\gamma,t} C_{s_{i-1}+\gamma,s_{j-1}+t_1p} .$$

In Δ_A put $a_1 = a_2 = \cdots = a_{s_1}$; $a_{s_1+1} = \cdots = a_{s_2-1}$ and so on, i.e. the a's which belong to C's which are in the same class B are put equal to one another. Add

the corresponding rows of Δ_A' then the elements in the row into which the sums are taken are also row elements of Δ_B. The order of $\Delta_B <$ order of Δ_A'. Hence certain columns of Δ_B are repeated in Δ_A'. Subtract the equal columns from one another and we reach a determinant in which Δ_B appears in one corner flanked by a zero matrix. Hence $|\Delta_A'| = |\Delta_B|\varphi$.

3. Factorization of the m-ary p-ic

The characteristic Δ_x is linear in the indeterminates x_i, integral in the c_{ijk} and satisfies the characteristic equation and the reduced characteristic equations which are homogeneous in the x_i's and integral in the c_{ijk}.

DEFINITION 1. A matrix in which the elements are linear in the m-indeterminates x_i and rational or integral in the coefficients of an m-ary p-ic over a field K of characteristic $\neq 2$ will be termed a *linear matrix* over $K(x_1, x_2, \cdots, x_m)$.

We proceed to factorize any m-ary p-ic over K as a product of p mutually commutative linear matrices.

THEOREM 10. *An m-ary quadratic can be expressed as the product of commutative linear matrices of order $\leqq 2^{m-1}$.*

Proceed by induction:

$$ax^2 + bxy + cdy^2 = \begin{bmatrix} ax + by, & cy \\ -dy, & x \end{bmatrix}\begin{bmatrix} x, & -cy \\ dy, & ax + by \end{bmatrix} \tag{7}$$

$$\begin{aligned} ax^2 + by^2 + cz^2 + dyz + ezx + fxy & \\ &= ax^2 + x(fy + ez) + by^2 + dyz + cz^2 \\ &= ax^2 + x(fy + ez) + C\cdot D \\ &= \begin{bmatrix} ax + fy + ez, & c \\ -D, & x \end{bmatrix}\cdot\begin{bmatrix} x, & -C \\ D, & ax + fy + ez \end{bmatrix} \end{aligned} \tag{8}$$

where

$$C = \begin{bmatrix} by + dz, & cz \\ -z, & y \end{bmatrix}, \qquad D = \begin{bmatrix} y, & -cz \\ z, & by + dz \end{bmatrix}$$

and where in (8) x, $ax + fy + ez$ are scalar matrices of order 2, i.e.

$$= \begin{bmatrix} ax + fy + ez, & 0, & by + dz, & cz \\ 0, & ax + fy + ez, & -z, & y \\ -y, & cz, & x, & 0 \\ -z, & -by - dz, & 0, & x \end{bmatrix}$$

$$\times \begin{bmatrix} x, & 0, & -by - dz, & -cz \\ 0, & x, & z, & -y \\ y, & -cz, & ax + fy + cz, & 0 \\ z, & by + dz, & 0, & ax + fy + ez \end{bmatrix}$$

Assume that the $(m - 1)$-ary quadratic can be expressed in commutative linear matrix factors of order 2^{m-2}. Then the m-ary quadratic can be written $f = ax^2 + bx + C - D$ where b is linear in $x_2, x_3, \cdots, x_m$ and C and D are commutative linear matrices of order 2^{m-2}. Hence

$$f = \begin{bmatrix} ax + b, & C \\ -D, & x \end{bmatrix} \begin{bmatrix} x, & -C \\ D, & ax + b \end{bmatrix}$$

is factorized in linear commutative matrices of order 2^{m-1}. Hence the theorem may be proved by induction.

In special cases the matrix order may be less than 2^{m-1} and if the restriction as to rationality be abandoned the matrix order may also be reduced further.

COROLLARY 1. *Similarly the bilinear form* $\sum a_{ij}x_iy_j$ *may be factorized in commutative linear matrices of order* $\leqq 2^m$.

COROLLARY 2. *A linear form* $\sum a_ix_i$ *may be factorized in commutative linear matrices of order* 2^{2m} *in which the elements are either* $\pm x_i$ *or* $\pm a_i$ *or zero.*

COROLLARY 3. *The determinant of any matrix factor* $= f^t$, $t = 2^{m-1}$.

COROLLARY 4. *The factor matrices satisfy the same reduced and the same characteristic equations.*

A *tower* of matrix rings will now be constructed in which the m-ary p-ic may be factorized in p commutative linear matrices.

Let $f = A_1A_2 \cdots A_\gamma + B_1B_2 \cdots B_s$ be scalar in $K(x_1, x_2, \cdots, x_m)$ and let $A_1, A_2, \cdots, A_\gamma$ be mutually commutative linear matrices of the same order μ and $B_1, B_2, \cdots, B_s$ be also mutually commutative linear matrices and of the same order ν. Let $A = \sum \alpha_iA_i$ and $B = \sum \beta_jB_j$, α_i, β_j being indeterminates. In the direct product the matrix rings in which A and B lie let $A \to A'$ where A is repeated ν times in the leading diagonal and let $B \to B'$ in which each element of B is repeated μ times as a scalar matrix. Then $A'B' = B'A' = A \times B$. Hence in the total matrix ring of order $\mu\nu$ every A' is commutative with every B_j' as well as with the remaining A_k''s. Also $(A_iA_j)' = A_i'A_j'$ and $(B_iB_j)' = B_i'B_j'$. Further A_i', B_j' satisfy the same reduced equations as A_i and B_j respectively. Suppose also that $A_1A_2 \cdots A_\gamma$ and $B_1B_2 \cdots B_s$ are each scalars in $K(x_1, x_2, \cdots, x_m)$ then so are $A_1'A_2' \cdots A_\gamma'$ and $B_1'B_2' \cdots B_s'$ in the new matrix ring. Hence $A_1'A_2' \cdots A_\gamma' + B_1'B_2' \cdots B_s'$ is a scalar matrix of order $\mu\nu$ in which f is repeated $\mu\nu$ times in the leading diagonal. In general $A_i'A_j' + B_l'B_m' \neq (A_iA_j + B_lB_m)'$.

We proceed to prove by induction the principal theorem:

THEOREM 11. *An m-ary p-ic, f, over a commutative ring K can be expressed as the product of p commutative linear matrices of the same order each linear and integral in the coefficients and, if regular in one of the indeterminates, having $f = 0$ as its reduced equation.*

Assume first that f is regular in x, i.e. $f = x^p + a_1x^{p-1} + a_2x^{p-2} + \cdots + a_p$ where a_i, $i = 1, 2, \cdots, p$ is of degree i and is homogeneous in $x_2, x_3, \cdots, x_p$ the coefficients being in the commutative ring K. For the purpose of factoriza-

tion it is unnecessary to take the coefficient of x^p to be 1 but it simplifies the discussion to assume this. Then

$$f = x[x^{p-1} + a_1x^{p-2} + \cdots + a_{p-1}] + a_p .$$

Assume that it has been proved that both the m-ary p-ic a_p and the m-ary $(p - 1)$-ic [ ] can both be expressed as products of commutative linear matrices viz:

$$f = x(x - x_2)(x - x_3) \cdots (x - x_p) + A_1A_2 \cdots A_p . \tag{9}$$

Write these as matrices x_i', A_i' in the direct product M of the matrix rings in which they lie so that, as above, x_i' and A_j' are mutually commutative and for convenience ignore the $'$, then in (9) we may regard the A's and x's as mutually commutative in the matrix ring $K(M)$. Then $x - u$ is a linear matrix factor of f where

$$\begin{bmatrix} x_2, & -A_1, & 0, 0, \cdots & 0, & 0 \\ 0, & x_3, & -A_2, 0, \cdots & 0, & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & 0, 0, \cdots & x_p, & -A_{p-1} \\ -A_p, & 0, & 0, 0, \cdots & 0, & 0 \end{bmatrix}. \tag{10}$$

The remaining factor, found by ordinary division in $K(M)$, is the homogeneous $(p - 1)$-ic in $x, x_2, \cdots, x_m, u$, viz. the $(m + 1)$-ary $(p - 1)$-ic in $K(M, x, x_2, \cdots, x_m, u)$

$$\begin{aligned} &x^{p-1} + (a_1 + u)x^{p-2} + (u^2 + a_1u + a_2)x^{p-3} \\ &\qquad + \cdots + (u^{p-1} + a_1u^{p-2} + \cdots + a_{p-1}). \end{aligned} \tag{11}$$

Hence if we also assume that the $(m + 1)$-ary $(p - 1)$-ic can be factorized as desired then so can the m-ary p-ic. The theorem is true for all the $(m + 1)$-ary quadratic. Hence to complete the induction it is necessary to prove it for the binary p-ic.

The binary p-ic may be factorized in commutative matrices of order $\leqq p!$ of the Galois group of $f = 0$ but such matrices are non-linear in the coefficients, their order not being large enough for our purpose.[3]

The m-ary cubic may be written as $f = x(m\text{-ary quadratic}) + (m - 1)$-ary cubic.

Assume that the theorem is true for the $(m - 1)$-ary cubic. Then, since it is true by theorem 10 for the m-ary quadratic, $f = x(x - x_2)(x = x_3) + A_1A_2A_3$. Write this in $K(M)$ as $f' = x(x - x_2')(x - x_3') + A_1'A_2'A_3'$ where $x_2' + x_3'$, $x_2'x_3'$ and $A_1'A_2'A_3'$ are scalars viz. $x_2 + x_3$, x_2x_3, $A_1A_2A_3$ respectively. Then $x - u$ is a linear factor

$$u = \begin{bmatrix} x_2', & -A_1', & 0 \\ 0, & x_3', & -A_2' \\ -A_3', & 0, & 0 \end{bmatrix}$$

[3] A. R. Richardson. Quart. Journal Math., 1934–35, pp. 256–270.

The remaining factor is

$$x^2 + x[u - (x_2' + x_3')] + [u^2 - u(x_2' + x_3') + x_2'x_3']$$

an $(m + 1)$-ary quadratic in $x, x_2, x_3, \cdots, x_m, u$ which can be factorized as desired. The theorem is true for the binary cubic; therefore, by induction, it is true for the m-ary cubic.

The actual factors are:

$$\begin{bmatrix} x - u, & 0 \\ 0\ , & x - u \end{bmatrix} \begin{bmatrix} x + u - (x_2' + x_3'), & u - x_2' \\ -u + x_3'\ , & x \end{bmatrix}$$

$$\begin{bmatrix} x\ , & -u + x_2' \\ u - x_3', & x + u - (x_2' + x_3') \end{bmatrix}$$

where $x + u - (x_2' + x_3')$ is interpreted as

$$\begin{bmatrix} x - x_3', & -A_1'\ , & 0 \\ 0\ , & x - x_2', & -A_2' \\ -A_3'\ , & 0\ , & x - (x_2' + x_3') \end{bmatrix}$$

and $u - x_2'$ as

$$\begin{bmatrix} 0\ , & -A_1'\ , & 0 \\ 0\ , & x_3' - x_2', & -A_2' \\ -A_3', & 0\ , & -x_2' \end{bmatrix}$$

Hence

$$f = \left[\begin{array}{ccc:ccc} x - x_2, & A_1\ , & 0 & & & \\ 0\ , & x - x_3, & A_2 & & & \\ A_3\ , & 0\ , & x & & & \\ \hdashline & & & x - x_2, & A_1\ , & 0 \\ & & & 0\ , & x - x_3, & A_2 \\ & & & A_3\ , & 0\ , & x \end{array}\right]$$

$$\times \left[\begin{array}{ccc:ccc} x - x_3\ , & -A_1\ , & 0 & 0\ , & -A_1\ , & 0 \\ 0\ , & x - x_2, & -A_2 & 0\ , & x_3 - x_2, & -A_2 \\ -A_3\ , & 0\ , & x - (x_2 + x_3) & -A_3, & 0\ , & -x_2 \\ \hdashline -x_2 + x_3, & A_1\ , & 0 & x\ , & 0\ , & 0 \\ 0\ , & 0\ , & A_2 & 0\ , & x\ , & 0 \\ A_3\ , & 0\ , & x_3 & 0\ , & 0\ , & x \end{array}\right]$$

$$\times \left[\begin{array}{ccc:ccc} x\ , & 0\ , & 0 & 0\ , & A_1\ , & 0 \\ 0\ , & x\ , & 0 & 0\ , & x_2 - x_3, & A_2 \\ 0\ , & 0\ , & x & x - x_3, & -A_1\ , & 0 \\ \hdashline x_2 - x_3, & -A_1, & 0 & x - x_3, & -A_1\ , & 0 \\ 0\ , & 0\ , & -A_2 & 0\ , & x - x_2, & -A_2 \\ -A_3\ , & 0\ , & -x_3 & -A_3\ , & 0\ , & x - (x_2 + x_3) \end{array}\right]$$

where $'$ has been dropped. It will be noted that each matrix *zero* is semi-reduced, that the first and the product of the second and third are completely reduced and that the product of the three factors is the scalar matrix having f in each element of the leading diagonal and of order $3!\mu\nu$.

The binary quadratic depends for its factorization on that of a ternary cubic which, as above, can be factorized. Hence the theorem is true for the binary quadratic. *The binary p-ic* depends for its factorization on that of the ternary $(p - 1)$-ic only since the term $a_p y^p$ is already in linear commutative factors. Hence, by induction, the theorem is true for the binary p-ic and, by further induction, for the m-ary p-ic.

COROLLARY 1. *The theorem holds for a multilinear form.*

COROLLARY 2. *Since the matrix factors are linear in the coefficients of the forms, $F = \sum \lambda_i f_i$ will have factors linear in the λ_i's. Therefore any set of m-ary p-ics in the same indeterminates may be factorized simultaneously in the same matrix ring.*

COROLLARY 3. *There are minimum orders to the matrix rings in which a special m-ary p-ic or in which every m-ary p-ic over K can be factorized linearly.*

COROLLARY 4. *The factorization is not unique but since the invariant factors and reduced equations are the same, two factorizations of the same order are similar.*

COROLLARY 5. *f may be regarded as a form in $m + p'$ indeterminates if the p' coefficients are themselves regarded as indeterminate. Hence f may be factorized in matrices in which the elements are sums of the coefficients and indeterminates each multiplied by $+1$, -1 or 0. Such a factorization will be termed complete.*

Factorization of irregular f. The preceding factorization applies to regular f's of which the characteristic and reduced equations of the class-ring are specimens. An irregular f may be factorized thus: arrange the terms of f in any order and consider two of them $\alpha_1\alpha_2 \cdots \alpha_p + \beta_1\beta_2 \cdots \beta_p$ where the α's and β's need not all be different. Then

$$u = \begin{bmatrix} \alpha_1, & \beta_1, & 0, \cdots, & 0 \\ 0, & \alpha_2, & \beta_2, \cdots, & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \beta_p, & 0, & 0, \cdots, & \alpha_p \end{bmatrix}$$

satisfies the equation

$$\prod (\lambda - \alpha_i) + (-1)^p \prod \beta_i = 0. \tag{12}$$

The conjugates of u may be calculated thus: divide by $\lambda - u$ obtaining a $(p - 1)$-ic regular in λ. This may be factorized as in (11). Then u and its conjugates are linear commutative matrices such that

$$uu_2 \cdots u_p = \alpha_1\alpha_2 \cdots \alpha_p + \beta_1\beta_2 \cdots \beta_p .$$

Repeat the process with a further term $\gamma_1\gamma_2 \cdots \gamma_p$ of f. The γ's are commutative with the u's and the above method leads to a set of conjugate commutative linear matrices $\omega\omega_2 \cdots \omega_p = \alpha_1\alpha_2 \cdots \alpha_p + \beta_1\beta_2 \cdots \beta_p + \gamma_1 \cdots \gamma_p$. By successive applications of the process, f itself may be factorized as desired. This is the practical method of procedure.

The matrix order in which the factorization takes place may be reduced if the restriction as to rationality in the coefficients is abandoned. Thus the quaternary cubic may be expressed *irrationally* as

$$f = x_1^3 + x_2^3 + x_3^3 + x_4^3 + x_5^3$$

$$x_2^3 + x_3^3 = \begin{bmatrix} x_2 + x_3, & 0 \\ 0, & x_2 + x_3 \end{bmatrix} \begin{bmatrix} x_2 - x_3, & x_3 \\ -x_3, & x_2 \end{bmatrix} \begin{bmatrix} x_2, & -x_3 \\ x_3, & x_2 - x_3 \end{bmatrix} = ABC$$

$$x_4^3 + x_5^3 = \begin{bmatrix} x_4 + x_5, & 0 \\ 0, & x_4 + x_5 \end{bmatrix} \begin{bmatrix} x_4 - x_5, & x_5 \\ -x_5, & x_4 \end{bmatrix} \begin{bmatrix} x_4, & -x_5 \\ x_5, & x_4 - x_5 \end{bmatrix} = PQR$$

Let

$$u = \begin{bmatrix} A & P & 0 \\ 0 & B & Q \\ R & 0 & C \end{bmatrix}$$

$$S \cdot T = \begin{bmatrix} u - (A + B + C), & 0, & B + C, & B \\ 0, & u - (A + B + C), & -C, & A \\ -A, & B, & u, & 0 \\ -C, & -B - C, & 0, & u \end{bmatrix}$$

$$\times \begin{bmatrix} u, & 0, & -B - C, & -B \\ 0, & u, & C, & -A \\ A, & -B, & u - (A + B + C), & 0 \\ C, & B + C, & 0, & u - (A + B + C) \end{bmatrix}$$

Then:

$$\begin{bmatrix} u, & 0 \\ 0, & u \end{bmatrix} \begin{bmatrix} u - (A + B + C), & S \\ -T, & 0 \end{bmatrix} \begin{bmatrix} 0, & -S \\ T, & u - (A + B + C) \end{bmatrix} = \bar{u}\bar{v}\bar{w}$$

for A, B, C, P, Q, R are mutually commutative.

$$f = x_1^3 + \bar{u}\bar{v}\bar{w} = \begin{bmatrix} x_1 - w, & 0 \\ 0, & x_1 - w \end{bmatrix} \begin{bmatrix} x_1 + w, & -w \\ w, & x_1 \end{bmatrix} \begin{bmatrix} x_1, & w \\ -w, & x_1 + w \end{bmatrix}$$

where

$$w = \begin{bmatrix} 0, & -\bar{u}, & 0 \\ 0, & 0, & -\bar{v} \\ -\bar{w}, & 0, & 0 \end{bmatrix}$$

Hence f may be factorized in matrices of order 144 which is the same as that for the ternary cubic having rational factors. If the ternary cubic is expressed irrationally as $x^3 + y^3 + z^3 + bxyz$ it can be factorized in linear matrices of order 48.

If f is regular in x then the reduced equation for each matrix factor is $f = 0$. Let

$$(13) \qquad u = \begin{bmatrix} A_1, & C_1, & 0, & \cdots, & 0, & 0 \\ 0, & A_2, & C_2, & \cdots, & 0, & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & 0, & \cdots, & A_{p-1}, & C_{p-1} \\ C_p, & 0, & 0, & \cdots, & 0, & A_p \end{bmatrix}$$

where A_i, C_j; $i, j = 1, 2, \cdots, p$ are mutually commutative matrices each of order μ. Let $T = (t_{i,j})$ where i, j denote row and column respectively and

$$t_{i,j} = (-1)^{p-i} c_1 c_2 \cdots c_{i-1} c_p h_{p-(i+j)}(A_1, A_2, \cdots, A_j, A_p),$$
$$j < p, \qquad t_{i,p} = (-A_p)^{p-1}$$

where $h_s(\alpha_1, \alpha_2 \cdots \alpha_j)$ is the sum of the homogeneous products of $\alpha_1, \alpha_2, \cdots \alpha_j$ taken s at a time and $h_0 = 1$, $h_s = 0$. Then

$$(14) \qquad TuT^{-1} = u' = \begin{bmatrix} a_1, & a_2, & \cdots, & a_{p-1}, & a_p \\ -1, & 0, & \cdots, & 0, & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & \cdots, & -1, & 0 \end{bmatrix}$$

where $a_p = A_1 A_2 \cdots A_p + (-1)^{p-1} c_1 c_2 \cdots c_p$ and a_i is the elementary symmetric function, $\sum A_1 A_2 \cdots A_i$ of $A_1, A_2, \cdots, A_p$ of degree i. Hence u' satisfies $f = 0$. In general T is non-singuar in $K(M)$ since $|T| = t_{p-1,1}\, t_{p-2,2} \cdots t_{p-i,i} t_{1,p-1} = (-1)^{\nu} c_1^{p-2} c_2^{p-3} \cdots c_{p-2} c_p^{p-1}$ where $\nu = 2$ or 1 according as $p \equiv 0$ or 1; or 2 or 3 (mod 4).

Hence unless all the C's are singular, T is non-singular in $K(M)$ and this suffices for our present purpose, for in the m-ary p-ic $a_p = A_1 A_2 \cdots A_p \neq 0$.

Each element in (14) is a scalar matrix of order μ. Hence u' is similar to u and can be transformed into the matrix in which the canomical elementary matrix in K

$$u'' = \begin{bmatrix} a_1, & a_2, & \cdots & , & a_p \\ -1, & 0, & \cdots & , & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & \cdots & -1, & 0 \end{bmatrix}$$

is repeated μ times in the leading diagonal. Hence u satisfies the same reduced equation in K as u'' viz. $f = 0$.

The $(p - 1)$ conjugates of $\breve{u}$ also satisfy $f = 0$ as reduced equation in $K(M)$. Thus u_2 satisfies (11) in $K(M, u)$ and since u_2 is equivalent to an elementary canonical matrix in $K(M, u)$, (11) is its reduced equation. Hence, in $K(M)$, $f = 0$ is its reduced equation. A repetition of the argument shows that u_3 is equivalent to an elementary canonical matrix having a reduced equation in $K(M, u, u_2)$ of degree $(p - 1)$ and hence having (11) as reduced equation in $K(M, u)$, i.e. $f = 0$ in $K(M)$. Similarly the other conjugates have $f = 0$ as reduced equation.

These conjugates of u can be transformed into (14). Actual transforming matrices for cubics and quadratics are:

$$T = \begin{bmatrix} x_3^2, & A_1(x_2 + x_3), & 0, & 0, & A_1 x_2, & A_1 A_2 \\ -x_3, & -A_1, & 0, & 0, & -A_1, & 0 \\ 1, & 0, & 0, & 0, & 0, & 0 \\ -2A_3 x_3, & -A_1 A_3, & -x_3(x_2 + x_3), & -A_1 x_3, & -A_1 x_3, & -x_2 x_3 \\ A_3, & 0, & x_3, & 0, & 0, & 0 \\ 0, & 0, & 0, & 0, & 0, & 1 \end{bmatrix}$$

transforms

$$\begin{bmatrix} \bar{u}, & 0 \\ 0, & \bar{u} \end{bmatrix}$$

into the second zero where

$$\bar{u} = \begin{bmatrix} -1, & a_2, & a_3 \\ a_1, & 0, & 0 \\ 0, & -1, & 0 \end{bmatrix}$$

In general T is non-singular, e.g. for the binary cubic

$$| T | = a_3(a_3^2 - a_1a_2).$$

For the quadratic

$$T = \begin{bmatrix} T_{11}, & T_{12}, & T_{13} \\ T_{21}, & T_{22}, & T_{23} \\ T_{31}, & T_{32}, & T_{33} \end{bmatrix}$$

where

$$T_{11} = \begin{bmatrix} -a_1a_2 + a_3, & a_2(a_1^2 - a_2), & -a_3(a_1^2 - a_2), & a_4(a_1^2 - a_2) \\ -a_2, & a_1a_2, & -a_1a_3, & a_1a_4 \\ 0, & a_2, & -a_3, & a_4 \\ 1, & 0, & 0, & 0 \end{bmatrix}$$

$$T_{12} = \begin{bmatrix} a_1a_3 - a_4, & a_2a_3 - a_1^2a_2, & a_4(a_1^2 - a_2), & 0 \\ a_3, & -a_1a_3, & a_1a_4, & 0 \\ 0, & -a_3, & a_4, & 0 \\ 0, & 0, & 0, & 0 \end{bmatrix}$$

$$T_{13} = \begin{bmatrix} -a_1a_4, & a_4(a_1^2 - a_2), & 0, & 0 \\ -a_4, & a_1a_4, & 0, & 0 \\ 0, & a_4, & 0, & 0 \\ 0, & 0, & 0, & 0 \end{bmatrix}$$

$$T_{21} = \begin{bmatrix} 0, & -1, & a_1, & 0 \\ 0, & 0, & 1, & 0 \\ 0, & 0, & 0, & 0 \\ 0, & 0, & 0, & 0 \end{bmatrix}, \quad T_{22} = \begin{bmatrix} -1, & a_1, & -a_2, & 0 \\ 0, & 0, & 0, & 0 \\ 0, & 0, & 1, & 0 \\ 0, & 0, & 0, & 0 \end{bmatrix}, \quad T_{23} = \begin{bmatrix} 0, & 0, & 0, & a_4 \\ 0, & 0, & 0, & 0 \\ 0, & 0, & 0, & 0 \\ 0, & 0, & 1, & 0 \end{bmatrix},$$

$$T_{31} = \begin{bmatrix} 0, & 0, & -1, & a_1 \\ 0, & 0, & 0, & 1 \\ 0, & 0, & 0, & 0 \\ 0, & 0, & 0, & 0 \end{bmatrix}, \quad T_{32} = \begin{bmatrix} 0, & -1, & a_1, & -a_2 \\ 0, & 0, & 0, & 0 \\ 0, & 0, & 0, & 1 \\ 0, & 0, & 0, & 0 \end{bmatrix}, \quad T_{33} = \begin{bmatrix} -1, & a_1, & -a_2, & a_3 \\ 0, & 0, & 0, & 0 \\ 0, & 0, & 0, & 0 \\ 0, & 0, & 0, & 1 \end{bmatrix}$$

transforms

$$\begin{bmatrix} u, & 0, & 0 \\ 0, & u, & 0 \\ 0, & 0, & u \end{bmatrix}$$

where

$$u = \begin{bmatrix} -a_1, & a_2, & -a_3, & a_4 \\ -1, & 0, & 0, & 0 \\ 0, & -1, & 0, & 0 \\ 0, & 0, & -1, & 0 \end{bmatrix}$$

into the second zero, viz:

$$\begin{bmatrix} -u - a_1, & u^2 + a_1 u + a_2, & -(u^3 + a_1 u^2 + a_2 u + a_3) \\ -1, & 0, & 0 \\ 0, & -1, & 0 \end{bmatrix}.$$

A result which is useful when calculating the conjugate zeros of u is: if

$$u = \begin{bmatrix} -a_1, & a_2, & -a_3, & \cdots, & (-1)^p a_p \\ -1, & 0, & 0, & \cdots, & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & \cdots & -1, & 0 \end{bmatrix}$$

then

$$u^s + a_1 u^{s-1} + a_2 u^{s-2} + \cdots + a_s$$

$$= \left[\begin{matrix} 0, & a_{s+1}, & -a_{s+2}, & \cdots & \cdots \\ 0, & 0, & a_{s+1}, & -a_{s+2}, & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & (-1)^{s-2} a_2, & \cdots & \cdots \\ (-1)^s, & (-1)^{s-1} a_1, & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & \cdots & \cdots, & \cdots \end{matrix}\right.$$

$$\left.\begin{matrix} +(-1)^s a_p, & 0, & 0, & \cdots & 0, & 0 \\ +(-1)^s a_{p-1}, & (-1)^s a_p, & 0, & \cdots & 0, & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0, & a_{s+1}, & -a_{s+2}, & \cdots & (-1)^{s-1} a_{p-1}, & (-1)^s a_p \\ a_s, & 0, & 0, & \cdots & 0, & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ (-1)^s, & (-1)^{s-1} a_1, & \cdots & \cdots & \cdots & a_s \end{matrix}\right]$$

and in particular

$$u^{p-1} + a_1 u^{p-2} + \cdots + a_{p-1} = (u - x_2)(u - x_3) \cdots (u - x_p)$$

$$= \begin{bmatrix} 0, & a_p, & 0, & \cdots & 0 \\ 0, & 0, & a_p, & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & \cdots & \cdots & a_p \\ (-1)^{p-1}, & (-1)^{p-2} a_1, & \cdots & & a_{p-1} \end{bmatrix}$$

4. Pseudo-representations

The characteristic polynomial of the class-ring may now be factorized in linear commutative matrix factors

$$|\lambda - \Delta_x| = \prod_j [\lambda - \sum x_i X_{i,j}]$$

$X_{i,j}$ being the matrix coefficient of x_i in the j^{th} linear factor. The number of factors is equal to the number of classes and each factor gives rise to a matrix-ring in which the correspondence $C_i \rightarrow X_{i,j}$ can be established.

DEFINITION. *$X_{i,j}$ will be termed a representative of C_i in the j^{th} matrix-ring.*

DEFINITION. *The p matrix-rings will be termed conjugate rings.*

In group theory the correspondence $C_i \rightarrow X_{i,j}/X_{0,j}$ is a representation, but in general this is not so. Even when $C_i \rightarrow X_{i,1}$ gives a representation, the conjugate matrices $X_{i,j}$ may not do so, although the general number in the j^{th} matrix ring is similar to the corresponding general number in the k^{th} ring, i.e. although $T_{j,k}$ exists such that $T_{j,k}^{-1} (\sum x_i X_{i,j})\ T_{j,k} = \sum x_i X_{i,k}$, $T_{j,k}$ depends on $x_i's$ and does not transform $X_{i,j}$ into $X_{i,k}$. There may however exist a *pseudo-representation* in which if $C_i C_j = \sum C_{ijp} C_p$ then

$$\prod_k [X_{i,k}\, X_{j,k} - \sum c_{ijp}\, X_{p,k}] = 0, \tag{15}$$

although the factors in this product are not necessarily conjugates.

Pseudo-representation of the total matrix algebra of two rowed matrices. Let $e_{11}, e_{12}, e_{21}, e_{22}$ denote $\begin{pmatrix}1, & 0\\ 0, & 0\end{pmatrix}, \begin{pmatrix}0, & 1\\ 0, & 0\end{pmatrix}, \begin{pmatrix}0, & 0\\ 1, & 0\end{pmatrix}, \begin{pmatrix}0, & 0\\ 0, & 1\end{pmatrix}$ respectively. The characteristic polynomial of the general number $x = \alpha e_{11} + \beta e_{12} + \gamma e_{21} + \delta e_{22}$ is

$$x^2 - x(\alpha + \delta) + \alpha\delta - \beta\gamma.$$

One factorization of this is:

$$\begin{bmatrix} x - (\alpha + \delta), & 0, & \alpha, & \beta \\ 0, & x - (\alpha + \delta), & \gamma, & \delta \\ -\delta, & \beta, & x, & 0 \\ \gamma, & -\alpha, & 0, & x \end{bmatrix}
\begin{bmatrix} x, & 0, & -\alpha, & -\beta \\ 0, & x, & -\gamma, & -\delta \\ \delta, & -\beta, & x - (\alpha + \beta), & 0 \\ -\gamma, & \alpha, & 0, & x - (\alpha + \beta) \end{bmatrix}$$

which gives rise to two sets of matrices $X_{i,j}$ corresponding respectively to e_{11}, e_{12}, e_{21}, e_{22} viz.:

$$\begin{bmatrix} 1, & 0, & -1, & 0 \\ 0, & 1, & 0, & 0 \\ 0, & 0, & 0, & 0 \\ 0, & 1, & 0, & 0 \end{bmatrix}, \begin{bmatrix} 0, & 0, & 0, & -1 \\ 0, & 0, & 0, & 0 \\ 0, & -1, & 0, & 0 \\ 0, & 0, & 0, & 0 \end{bmatrix},$$

$$\begin{bmatrix} 0, & 0, & 0, & 0 \\ 0, & 0, & -1, & 0 \\ 0, & 0, & 0, & 0 \\ -1, & 0, & 0, & 0 \end{bmatrix}, \begin{bmatrix} 1, & 0, & 0, & 0 \\ 0, & 1, & 0, & -1 \\ 1, & 0, & 0, & 0 \\ 0, & 0, & 0, & 0 \end{bmatrix} \tag{16}$$

$$(17)\quad \begin{bmatrix} 0, & 0, & 1, & 0 \\ 0, & 0, & 0, & 0 \\ 0, & 0, & 1, & 0 \\ 0, & -1, & 0, & 1 \end{bmatrix}, \begin{bmatrix} 0, & 0, & 0, & 1 \\ 0, & 0, & 0, & 0 \\ 0, & 1, & 0, & 0 \\ 0, & 0, & 0, & 0 \end{bmatrix}, \begin{bmatrix} 0, & 0, & 0, & 0 \\ 0, & 0, & 1, & 0 \\ 0, & 0, & 0, & 0 \\ 1, & 0, & 0, & 0 \end{bmatrix}, \begin{bmatrix} 0, & 0, & 0, & 0 \\ 0, & 0, & 0, & 1 \\ -1, & 0, & 1, & 0 \\ 0, & 0, & 0, & 1 \end{bmatrix}$$

These give a pseudo-representation for, using (16)

$$Z_1 = [xx' - \{(\alpha\alpha' + \beta\gamma')e_{11} + (\alpha\beta' + \beta\delta')e_{12} + \gamma\alpha' + \delta\gamma')e_{21} + (\gamma\beta' + \delta\delta')e_{22}\}]$$

$$= \begin{bmatrix} \delta\alpha' - \gamma\beta', & \alpha\beta' - \beta\alpha', & \beta\gamma' - \delta\alpha', & \beta\delta' - \delta\beta' \\ \delta\gamma' - \gamma\delta', & \alpha\delta' - \beta\gamma', & \gamma\alpha' - \alpha\gamma', & \gamma\beta' - \alpha\delta' \\ \delta\alpha' - \gamma\beta', & \alpha\beta' - \beta\alpha', & \beta\gamma' - \delta\alpha', & \beta\delta' - \delta\beta' \\ \delta\gamma' - \gamma\delta', & \alpha\delta' - \beta\gamma', & \gamma\alpha' - \alpha\gamma', & \gamma\beta' - \alpha\delta' \end{bmatrix}$$

and, using (17)

$$Z_2 = \begin{bmatrix} \beta\gamma' - \alpha\delta', & \alpha\beta' - \alpha'\beta, & \alpha\delta' - \beta\gamma', & \beta\alpha' - \alpha\beta' \\ \delta\gamma' - \gamma\delta', & \gamma\beta' - \delta\alpha', & \gamma\delta' - \delta\gamma', & \delta\alpha' - \gamma\beta' \\ \gamma\beta' - \alpha\delta', & \delta\beta' - \beta\delta', & \alpha\delta' - \gamma\beta', & \beta\delta' - \delta\beta' \\ \alpha\gamma' - \gamma\alpha', & \beta\gamma' - \delta\alpha', & \gamma\alpha' - \alpha\gamma', & \delta\alpha' - \beta\gamma' \end{bmatrix}$$

and $Z_1Z_2 = 0$, i.e. the matrices (16) and (17) give a pseudo-representation of the total matrix algebra. Nevertheless Z_2 is not a conjugate of Z_1.

There are two other matters which may be mentioned. In the group ring the characters are orthogonal. In the general case this is not so and a biorthogonal relation takes its place. Suppose that the matrix equations

$$\sum_i \alpha_{i,j} X_{i,j} = 1, \qquad \sum_j \alpha_{i,j} X_{i,s} = 0, \qquad s \neq j$$

can be solved for the $\alpha_{i,j}$. This will be so if the determinant $| X_{i,j} |$ regarded as a determinant in K is non-zero. Then the α's and X's are biorthogonal, viz:

$$\sum_j X_{i,j} \alpha_{i,j} = n; \qquad \sum_j X_{i,j} \alpha_{k,j} = 0;$$

$$\sum_i \alpha_{i,j} X_{i,j} = n; \qquad \sum_i \alpha_{i,j} X_{i,k} = 0,$$

where n is an element of K invariant in the class-ring; in the group ring it is the group order.

The matrix ring generated by the α's is not in general the same as that generated by the $X_{i,j}$ but the two rings and the corresponding polynomials are related in a way which has some geometrical significance.

The second matter of interest is that $f = 0$ is the reduced characteristic equation of each of the matrix factors. If $f = 0$ is reducible in K, e.g. $f = f_1 f_2 \cdots f_s$

then the matrices may be constructed, as above, having each of these factors or any combination of them as reduced equations. These remarks will be illustrated by examples.

EXAMPLE 1.

$$\begin{array}{c|cc} & C_1 & C_2 \\ \hline C_1 & C_1 & C_2 \\ C_2 & C_2 & C_1 + C_2 \end{array}$$

The characteristic equation is

$$x^2 - x(2x_1 + x_2) + (x_1^2 + x_1x_2 - x_2^2) = 0 \tag{18}$$

This may be factorized as

$$\begin{bmatrix} x - (x_1 + x_2), & x_2 \\ x_2 \quad , & x - x_1 \end{bmatrix} \begin{bmatrix} x - x_1, & -x_2 \\ -x_2 \ , & x - (x_1 + x_2) \end{bmatrix}$$

giving the true representations

$$\begin{pmatrix} 1, & 0 \\ 0, & 1 \end{pmatrix}, \begin{pmatrix} 0, & 1 \\ 1, & 1 \end{pmatrix}; \begin{pmatrix} 1, & 0 \\ 0, & 1 \end{pmatrix}, \begin{pmatrix} 1\ , & -1 \\ -1, & 0 \end{pmatrix}; = x_{11}, x_{12}; x_{21}, x_{22}$$

respectively. The corresponding $\alpha_{i,\,j}$ are

$$\begin{pmatrix} 3\ , & -1 \\ -1, & 2 \end{pmatrix}, \begin{pmatrix} -1, & 2 \\ 2\ , & 1 \end{pmatrix}; \begin{pmatrix} 2, & 1 \\ 1, & 3 \end{pmatrix}, \begin{pmatrix} 1\ , & -2 \\ -2, & -1 \end{pmatrix}; = \alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}$$

respectively. The biorthogonal relations are:

$$\begin{aligned} x_{11}\alpha_{11} + x_{12}\alpha_{12} &= 5; & x_{21}\alpha_{11} + x_{22}\alpha_{12} &= 0; \\ x_{11}\alpha_{21} + x_{12}\alpha_{22} &= 0; & x_{21}\alpha_{21} + x_{22}\alpha_{22} &= 5; \\ \alpha_{11}x_{11} + \alpha_{21}x_{21} &= 5; & \alpha_{11}x_{12} + \alpha_{21}x_{22} &= 0; \\ \alpha_{12}x_{11} + \alpha_{22}x_{21} &= 0; & \alpha_{12}x_{12} + \alpha_{22}x_{22} &= 5. \end{aligned}$$

A different factorization of (18)

$$\begin{bmatrix} x - (2x_1 + x_2), & 0 \quad , & x_1 + x_2, & x_2 \\ 0 \quad , & x - (2x_1 + x_2), & x_2 \quad , & x_1 \\ -x_1 \quad , & x_2 \quad , & x \quad , & 0 \\ x_2 \quad , & -x_1 - x_2 \quad , & 0 \quad , & x \end{bmatrix}$$

$$\begin{bmatrix} x\ , & 0 \ , & -x_1 - x_2 \quad , & -x_2 \\ 0\ , & x \ , & -x_2 \quad , & -x_1 \\ x\ , & -x_2\ , & x - (2x_1 + x_2), & 0 \\ -x_2, & x_1 + x_2, & 0 \quad , & x - (2x_1 + x_2) \end{bmatrix}$$

gives the pseudo-representations

$$\begin{bmatrix} 2, & 0, & -1, & 0 \\ 0, & 2, & 0, & -1 \\ 1, & 0, & 0, & 0 \\ 0, & 1, & 0, & 0 \end{bmatrix}, \begin{bmatrix} 1, & 0, & -1, & -1 \\ 0, & 1, & -1, & 0 \\ 0, & -1, & 0, & 0 \\ -1, & 1, & 0, & 0 \end{bmatrix};$$

$$\begin{bmatrix} 0, & 0, & 1, & 0 \\ 0, & 0, & 0, & 1 \\ -1, & 0, & 2, & 0 \\ 0, & -1, & 0, & 2 \end{bmatrix}, \begin{bmatrix} 0, & 0, & 1, & 1 \\ 0, & 0, & 1, & 0 \\ 0, & 1, & 1, & 0 \\ 1, & -1, & 0, & 1 \end{bmatrix} \equiv x_{11}, x_{12}; x_{21}, x_{22}$$

for which it is readily verified that

$$(x_{12}^2 - x_{11} - x_{12})(x_{22}^2 - x_{21} - x_{22}) = 0; \qquad (x_{11}^2 - x_{11})(x_{21}^2 - x_{21}) = 0$$

$$(x_{11}x_{12} - x_{12})(x_{21}x_{22} - x_{22}) = 0.$$

Example 2. *The symmetric group of order* 3.

The reduced equation of the group-ring is

$$[x - (x_1 + 3x_2 + 2x_3)][x - (x_1 - x_3)][x - (x_1 - 3x_2 + 2x_3] = 0$$

Taking the first two factors together and factorizing in matrices of order 2 we get as pseudo-representations

$$\begin{pmatrix} 2, & -1 \\ 1, & 0 \end{pmatrix}, \begin{pmatrix} -3, & 0 \\ -3, & 0 \end{pmatrix}, \begin{pmatrix} 1, & 1 \\ 2, & 0 \end{pmatrix}; \begin{pmatrix} 0, & 1 \\ -1, & 2 \end{pmatrix}, \begin{pmatrix} 0, & 0 \\ 3, & -3 \end{pmatrix}, \begin{pmatrix} 0, & -1 \\ -2, & 1 \end{pmatrix}$$

$$= x_{21}, x_{22}, x_{23}; x_{31}, x_{32}, x_{33}$$

representatives of the classes (1), (12), (123). Writing the remaining factors as

$$\begin{pmatrix} 1, & 0 \\ 0, & 1 \end{pmatrix}, \begin{pmatrix} 3, & 0 \\ 0, & 3 \end{pmatrix}, \begin{pmatrix} 2, & 0 \\ 0, & 2 \end{pmatrix} = x_{11}, x_{12}, x_{13}$$

the biorthogonal set is:

$$\begin{pmatrix} 1, & 0 \\ 0, & 1 \end{pmatrix}, \begin{pmatrix} 1, & 0 \\ 0, & 1 \end{pmatrix}, \begin{pmatrix} 1, & 0 \\ 0, & 1 \end{pmatrix}; \begin{pmatrix} 1, & 0 \\ -3, & 4 \end{pmatrix}, \begin{pmatrix} -1, & 0 \\ -1, & 0 \end{pmatrix}, \begin{pmatrix} 1, & 0 \\ 3, & -2 \end{pmatrix};$$

$$\begin{pmatrix} 1, & 3 \\ 0, & 4 \end{pmatrix}, \begin{pmatrix} -1, & 1 \\ 0, & 0 \end{pmatrix}, \begin{pmatrix} 1, & -3 \\ 0, & -2 \end{pmatrix}$$

with

$$\sum_j x_{i,j}\alpha_{ij} = 3!, \qquad \sum_j x_{ij}\alpha_{k,j} = 0; \qquad \sum_i \alpha_{ij}x_{ij} = 3!, \qquad \sum_i \alpha_{i,j}x_{i,k} = 0.$$

Similarly any other grouping of the factors of the characteristic equation leads to a pseudo-representation.

The above pseudo-representations generate total matrix rings.

We may also repeat one of the factors, e.g. corresponding to $[x - (x_1 - x_3)]^2$ we have the pseudo-representations

$$\begin{pmatrix} 2, & -1 \\ 1, & 0 \end{pmatrix}, \begin{pmatrix} 0, & 0 \\ 0, & 0 \end{pmatrix}, \begin{pmatrix} -2, & 1 \\ -1, & 0 \end{pmatrix}; \begin{pmatrix} 0, & 1 \\ -1, & 2 \end{pmatrix}, \begin{pmatrix} 0, & 0 \\ 0, & 0 \end{pmatrix}, \begin{pmatrix} 0, & -1 \\ 1, & -2 \end{pmatrix}$$

These do not generate total matrix rings.

EXAMPLE 3. *The symmetric group of order* 4.

Denote the classes (1), (12), (123), (12) (34), (1234) by C_α, C_β, C_γ, C_δ, C_ϵ respectively. Then the characteristic equation has as zeros

$$\alpha + 6\beta + 8\gamma + 3\delta + 6\epsilon; \quad \alpha - 6\beta + 8\gamma + 3\delta - 6\epsilon; \quad \alpha - 4\gamma + 3\delta;$$
$$\alpha + 2\beta - \delta - 2\epsilon; \quad \alpha - 2\beta - \delta + 2\epsilon.$$

Any combination of these gives a pseudo-representation, e.g. $(\alpha - \delta)^2 - 4(\beta - \epsilon)^2$ gives

$$\begin{pmatrix} 2, & -1 \\ 1, & 0 \end{pmatrix}, \begin{pmatrix} 0, & -2 \\ -2, & 0 \end{pmatrix}, \begin{pmatrix} 0, & 0 \\ 0, & 0 \end{pmatrix}, \begin{pmatrix} -2, & 1 \\ -1, & 0 \end{pmatrix}, \begin{pmatrix} 0, & 2 \\ 2, & 0 \end{pmatrix};$$
$$\begin{pmatrix} 0, & 1 \\ -1, & 2 \end{pmatrix}, \begin{pmatrix} 0, & 2 \\ 2, & 0 \end{pmatrix}, \begin{pmatrix} 0, & 0 \\ 0, & 0 \end{pmatrix}, \begin{pmatrix} 0, & -1 \\ 1, & -2 \end{pmatrix}, \begin{pmatrix} 0, & -2 \\ -2, & 0 \end{pmatrix}.$$

These generate simply isomorphic matrix-rings.

EXAMPLE 4. *The following set has the same representation as the symmetric group of order* 2.

$$\sum : \begin{array}{c|ccc} & a & b & c \\ \hline a & a & c & b \\ b & c & a & a \\ c & b & a & a \end{array}; \qquad A = a; \qquad C_1 = a, \qquad C_2 = b,c \qquad \text{mod } R_A.$$

Then

$$\begin{array}{c|cc} & C_1 & C_2 \\ \hline C_1 & C_1 & C_2 \\ C_2 & C_2 & 4C_1 \end{array}$$

and the characteristic equation is

$$(x - x_1 + 2x_2)(x - x_1 - 2x_2) = 0.$$

Hence dividing the representatives by the class orders 1, 2, respectively we get 1, 1, 1; 1, -1, -1, as representatives of a, b, c respectively. The class-ring is simply isomorphic with the group-ring generated from the symmetric group of order 2! in which $C_1 \to 1$, $C_2 \to 2(1, 2)$. Therefore the existence of a set of numbers having all the properties of group-characters cannot be taken as evidence of the existence of a group having these as characters.

Conclusion

In addition to the analogy stressed above between the group characteristics and the linear matrix representatives of classes found by factorization of the characteristic polynomial of the class-ring there are many other fields in which the factorization of the m-ary p-ic may be applied.

For example the theory of quadratic forms is known to depend on that of generalized quaternion algebras. This becomes apparent as soon as the form is factorized. There are also applications to the theories of algebraic functions, invariants and arithmetic.

ANNALS OF MATHEMATICS
Vol. 44, No. 1, January, 1943

ON HOMOTOPY TYPE AND DEFORMATION RETRACTS

BY R. H. FOX

(Received October 1, 1942)

It has recently been shown by J. H. C. Whitehead[1] that two complexes X and Y belong to the same homotopy type[2] if and only if there is a third complex W of which both X and Y are deformation retracts.[3] I shall show that this theorem holds not merely for complexes but for the most general spaces for which continuity has a meaning. The proof which I give is direct and constructive and avoids the extraneous notions of relative homology and relative homotopy groups which complicate Whitehead's proof.

The concept of homotopy type splits naturally into two concepts which I shall call right- and left-homotopy inversion. In theorems 3.3 and 3.4 I show that right-and-left inversion correspond respectively to deformation and retraction, thus replacing Whitehead's theorem by two "component" theorems. The necessary preliminary study of deformation, retraction, and inversion is carried out in §§1 and 2, and the mapping cylinder, the fundamental tool of our theory, is defined in §3. It should be noted that Whitehead's definition[4] of mapping cylinder is not really satisfactory for the general spaces considered here.

The fundamental theorems of this paper are theorems 3.1 and 3.2. They are generalizations of the theorems (3.3 and 3.4) discussed above. In §4 these fundamental theorems are applied (in another direction) to the Hopf-Pannwitz deformation and also to yield a new characterization of the closure of a homogenous n-dimensional polyhedron.

These theorems (3.1 and 3.2) are of considerable interest in themselves. They exhibit a duality which is quite striking and seem to indicate a relatively unexplored region which I might designate as "algebra of mapping classes". In this connection they should be compared with the fundamental theorem of fibre spaces[5] to which they bear an evident analogy.

In §§5 and 6 certain specializations are considered. They are to be regarded as trends in the following two directions (a) bridging the gap between homotopy type and nucleus[6] (b) bridging the gap between homotopy type and topological type. In §7 I develop an n-dimensional analogue of §3. This is in line with

[1] J. H. C. Whitehead, *Simplicial spaces, nuclei, and m-groups.* Proc. London Math. Soc. 45 (1939), 243–327. The proof referred to is on p. 278.

[2] W. Hurewicz, *Topologie der Deformationen* III. Proc. Akad. Amsterdam 39 (1936), p. 124.

[3] K. Borsuk, *Zur Kombinatorischen Eigenschaften der Retracte.* Fund. Math. 21 (1933), p. 91.

[4] J. H. C. Whitehead, loc. cit., p. 259.

[5] W. Hurewicz and N. E. Steenrod, *Homotopy Relations in Fibre Spaces*, Proc. Nat. Acad. 27 (1941), p. 62, theorem 1.

[6] J. H. C. Whitehead, loc. cit., p. 247.

the viewpoint of my Thesis,[7] especially §§2 and 13; analogous definitions and theorems using homology and n-dimensional homology are quite obvious and are omitted. The n-dimensional homotopy was selected because the n-dimensional homotopy type seems to be related to the so-called $(n + 1)$-group[6] of Whitehead in much the same way that homotopy type is related to nucleus.

1. Deformation and retraction

Mappings[8] f and g of a space A into a space D are said to be *homotopic* (notation: $f \simeq g$) if there is a mapping ξ (called a *homotopy* between f and g) of the product $A \times [0, 1]$ of A with the closed interval $0 \leqq t \leqq 1$ into D such that $\xi_0(a) = f(a)$ and $\xi_1(a) = g(a)$ for every $a \in A$. If A is a subset of D and f is the identity, so that $f(a) = a$, the homotopy ξ is called a *deformation* and the set A is said to be *deformable* in D into $g(A)$. If $D = A$ we say merely that A can be deformed into $g(A)$.

(1.1)[9] *If A can be deformed into B then a mapping f of A into itself is homotopic to the identity if (and only if) $f \mid B$*[10] *is homotopic to the identity.*

If ξ is a deformation of A into B and η is a homotopy between $f \mid B$ and the identity then a homotopy ζ between f and the identity is defined by

$$\begin{aligned} \zeta_t(a) &= \xi_{3t}(a), & 0 \leqq t \leqq \tfrac{1}{3}, \\ &= \eta_{2-3t}(\xi_1(a)), & \tfrac{1}{3} \leqq t \leqq \tfrac{2}{3}, \\ &= f(\xi_{3-3t}(a)), & \tfrac{2}{3} \leqq t \leqq 1. \end{aligned}$$

A subset B of a space A is said to be a *retract* of A if there is a mapping r (called a *retraction*) of A into B such that $r \mid B$ is the identity mapping of B. If r is homotopic to the identity mapping of A then B is called a *deformation retract*[3] of A, r is called a *deformation retraction* and the homotopy is called a *retracting deformation.*

THEOREM 1.2. *In order that B be a deformation retract of A it is necessary and sufficient that B be a retract of A and A be deformable into B.*

This follows from (1.1) by specializing f to be a retraction of A into B.

The example of a point B contained in an n-sphere A ($n \geqq 0$) shows that B may be a retract of A without A being deformable into B. The example of an n-sphere B contained in an $(n + 1)$-cell A ($n \geqq 0$) shows that A may be deformable into B without B being a retract of A. These two statements are equivalent to each other and to the Brouwer fixed point theorem.[11]

---

[7] R. H. Fox, *On the Lusternik-Schnirelmann Category*, Annals of Math. 42 (1941), 333–370.

[8] A mapping of a space A into a space D means a continuous function defined on A with values in D.

[9] This lemma is a trivial generalization of Satz IVa Hilfsatz, Alexandroff and Hopf, *Topologie*, p. 251.

[10] The partial mapping $f \mid B$ is the mapping of B defined by the rule $\{f \mid B\}(b) = f(b)$ for every $b \in B$.

[11] See W. Hurewicz and H. Wallman, *Dimension Theory*, Princeton Mathematical Series 4 (1941) Chapter 5, §1.

(1.3) *If ξ is a deformation of an ANR-set*[12] *A and B is the set of fixed points of ξ_1 then there is a homotopy ζ between $\xi_0 = 1$ and ξ_1 such that the points of B are fixed under each of the mappings ζ_t, $0 \leqq t \leqq 1$.*

Let η be the map of $(A \times [0] + B \times [0, 1] + A \times [1]) \times [0, 1]$ into A defined by

$$\begin{aligned} \eta_u(a, 0) &= a , && 0 \leqq u \leqq 1, && a \in A, \\ \eta_u(a, t) &= \begin{cases} \xi_{t/u}(a), & 0 \leqq t \leqq u \leqq 1, \quad a \in B, \\ a , & 0 \leqq u \leqq t \leqq 1, \quad a \in B, \end{cases} \\ \eta_u(a, 1) &= \xi_1(a) , && 0 \leqq u \leqq 1, && a \in A. \end{aligned}$$

Since ξ is an extension[13] to $A \times [0, 1]$ of the map η_1, and since A is an ANR-set, it follows that there is an extension[14] ζ to $A \times [0, 1]$ of η_0. This deformation ζ has the required properties.

THEOREM 1.4. *If A and a subset B are ANR-sets then either of the conditions* (i), (ii) *is necessary and sufficient for B to be a deformation retract of A.*

(i) *There is a retracting deformation ξ of A onto B such that the points of B are fixed under each of the mappings ξ_t, $0 \leqq t \leqq 1$.*

(ii) *There is a deformation ξ of A into B such that $\xi_t(B) \subset B$ for every $0 \leqq t \leqq 1$.*

The sufficiency of (i) and the implication (i) $\rightarrow$ (ii) are trivial. That condition (i) is necessary follows from (1.3) by imposing the condition $\xi_1(A) = B$ so that ξ becomes a retracting deformation. If (ii) is assumed then $\xi_1 \mid B$ and $\xi_0 \mid B$ are homotopic in B. Since B is a compact ANR-set and since ξ_1 is an extension of $\xi_1 \mid B$ to A (with values in B) it follows[14] that there is an extension r of $\xi_0 \mid B$ to A. This extension r is clearly a retraction of A onto B. By (1.2) it follows that B is a deformation retract of A.

Appreciation of theorem 1.4 is facilitated by consideration of several examples. The first example is due to Hopf and Pannwitz.[15] B is the pseudomanifold obtained by pinching a meridian of a torus to a point; A is obtained from B by spanning an equator with a 2-cell. A can be deformed into B but condition (ii) is not satisfied (hence B is not a deformation retract and condition (i) is not satisfied either). In the second example A is a bounded portion of the Cartesian plane and B is the set

$$\{0 \leqq x \leqq 1; y = 0\} + \textstyle\sum_{n=1}^{\infty} \{x = 1/n; 0 \leqq y \leqq 1\} + \{x = 0; 0 \leqq y \leqq 1\},$$

hence not an ANR. Condition (ii) is satisfied but B is not a retract of A, hence not a deformation retract of A. In the third example A is the set B of the previous example and B is the point $(0, 1)$. B is a deformation retract of A but

---

[12] R. H. Fox, *A Characterization of Absolute Neighborhood Retracts*, Bull. Am. Math. Soc. 48 (1942), 271–275.

[13] If f is a mapping of X into Y an extension f^* of f to a space $X^* \supset X$ is a mapping of X^* into Y such that $f^* \mid X = f$.

[14] The Borsuk-Kuratowski theorem: Fox, ibid, p. 273, and Dowker's proof of Borsuk's theorem: Hurewicz and Wallman, ibid, p. 86.

[15] Alexandroff and Hopf, loc. cit. p. 287, fig. 23.

condition (ii) is not satisfied.[16] In the fourth example A is as in the last example and B is the line segment $\{x = 0; 0 \leqq y \leqq 1\}$. Here B is a deformation retract of A. Condition (ii) is satisfied but not[16] condition (i).

2. Homotopy type

Two spaces A and B are said to belong to the same *homotopy type*[2] if there are mappings f of A into B and g of B into A such that the maps gf of A into itself and fg of B into itself are each homotopic to the identity (in A and B respectively). Belonging to the same homotopy type is an equivalence relation.[2]

If mappings f of A into B and g of B into A are such that $gf \simeq 1$ then g will be called a *left homotopy inverse* of f or, briefly, a *left inverse* of f and f will be called a *right inverse* of g. *A two-sided inverse* of f is a mapping which is both a right and left inverse of f. Thus A and B belong to the same homotopy type if and only if there is a mapping of A into B which has a two-sided inverse.

THEOREM 2.[17] *If f has both right and left inverses then it has a 2-sided inverse.*

Let g' and g'' be left and right inverses respectively and let $g = g'fg''$. Then $gf = g'fg''f \simeq g' 1f = g'f \simeq 1$ and $fg = fg'fg'' \simeq f1g'' = fg'' \simeq 1$.

It may happen, even for compact ANR-sets A and B, that no map of A into B has a 2-sided inverse although maps of A into B can be found with either right or left inverses. Let O denote the 1-sphere and 8 the figure-eight graph. Let $A = 8 \times 8 \times 8 \times \cdots$ and let $B = O \times 8 \times 8 \times \cdots = O \times A$, so that A and B are compact ANR-sets.[18] The map f' of A into B defined by $f'(t_1, t_2, \cdots) = (p, t_1, t_2, \cdots)$, $p \in O$, has a left inverse g' defined by $g'(t_1, t_2, \cdots) = (t_2, t_3, \cdots)$. The map f'' of A into B defined by $f''(t_1, t_2, t_3, \cdots) = (\alpha(t_1), t_2, t_3, \cdots)$, where α maps 8 into O by folding the top down over the bottom, has a right inverse g'' defined by $g''(t_1, t_2, t_3, \cdots) = (\beta(t_1), t_2, t_3, \cdots)$ where β maps O homeomorphically onto the bottom half of 8. The fundamental group $\pi_1(A)$ of A is the infinite direct product $F_2 \times F_2 \times F_2 \times \cdots$ and the fundamental group $\pi_1(B)$ of B is the direct product $F_1 \times F_2 \times F_2 \times \cdots = F_1 \times \pi_1(A)$, where F_i denotes the free group on i generators ($i = 1, 2$). Since $\pi_1(B)$ has an element $(a, 1, 1, \cdots)$ which commutes with every element of $\pi_1(B)$, and $\pi_1(A)$ has no such element, $\pi_1(A)$ and $\pi_1(B)$ are not isomorphic. Hence[19] no map of A into B can have a 2-sided inverse.

3. Mapping cylinder

For any mapping α of a space[20] M into a space N let $N + C_\alpha$ denote the space obtained from the space N and the cylinder $M \times [0, 1]$ by identifying the point

---

[16] R. H. Fox, *On the Lusternik-Schnirelmann Category*, Annals of Math. 42 (1941), p. 362.

[17] This proof of the theorem was shown to me by M. M. Day who has proved a theorem on partially ordered sets by exactly the same method.

[18] N. Aronszahn and K. Borsuk, *Sur la somme et le produit combinatoire des rêtractes absolus*. Fund. Math. 18 (1932), Theorem 6, p. 197.

[19] W. Hurewicz, *Topologie der Deformationen* III, Proc. Akad. Amsterdam 39 (1936), p. 125.

[20] Unless otherwise specified spaces considered at the same time are mutually separated.

$n \varepsilon N$ and the closed set $(\alpha^{-1}(n), 1) \varepsilon M \times [1]$. Precisely, $N + C_\alpha$ is the hyperspace[21] of the decomposition of $N + M \times [0, 1]$ into the points (m, t), $0 \leqq t < 1$, of $M \times [0, 1)$ and the (closed) sets $n + (\alpha^{-1}(n), 1)$ of $N + M \times [1]$. Denoting the identification mapping by i, it can easily be proved that $i \mid M$ and $i \mid N$ are homeomorphisms and that $i \mid (f(M) + M \times [0, 1])$ is the identification mapping of the induced decomposition of $f(M) + M \times [0, 1]$. (Note that $i \mid M \times [0, 1]$ is not necessarily the identification mapping of the induced decomposition of $M \times [0, 1]$). Accordingly denote by C_α the hyperspace of the induced decomposition of $f(M) + M \times [0, 1]$ and consider C_α, M and N as subsets of $N + C_\alpha$ so that $i \mid (M + N)$ is the identity mapping. (This justifies the notation $N + C_\alpha$). I shall call $N + C_\alpha$ the mapping cylinder[4] of α; the symbol $\langle m, t\rangle$ will denote the point $i(m, t)$ of C_α, so that $m = \langle m, 0\rangle$ and $\alpha(m) = \langle m, 1\rangle$. If M and N are compact metric then so also is $N + C_\alpha$.[22] The dimension of $N + C_\alpha$ is max $\{\dim M, 1 + \dim N\}$.[23]

THEOREM 3.1. *Let X, Y and Z be topological spaces and let θ be a mapping of X into Z. If f is a mapping of X into Y then there is a mapping g of Y into Z satisfying $gf \simeq \theta$ if and only if θ can be extended*[13] *to $Y + C_f$.*

If θ^* is an extension of θ to $Y + C_f$ let $g = \theta^* \mid Y$. The homotopy ξ defined by

$$\xi_t(x) = \theta^*(\langle x, t\rangle), \qquad x \epsilon X, \qquad 0 \leqq t \leqq 1,$$

is a homotopy between $\xi_0 = \theta^* \mid X = \theta$ and $\xi_1 = \theta^* f = gf$.

Suppose, conversely, that a map g satisfying $gf \simeq \theta$ has been given and that ξ is a homotopy between θ and gf. Then the map θ^* of $Y + C_f$ into Z defined by

$$\theta^*(\langle x, t\rangle) = \xi_t(x), \qquad \langle x, t\rangle \epsilon C_f,$$

$$\theta^*(y) = g(y), \qquad y \epsilon Y,$$

is an extension of θ to $Y + C_f$. $\theta^* \mid C_f$ is continuous because if V is an open set of Z then $f^{-1}(V)$ is an open set of $X \times [0, 1]$ which is the union of sets of the decomposition $\{i^{-1}(\langle x, t\rangle)\}$.

THEOREM 3.2. *Let X, Y, Z, and θ be as in theorem 3.1. If g is a mapping of Y into Z then there is a mapping f of X into Y satisfying $gf \simeq \theta$ if and only if θ is homotopic in $Z + C_g$ to a map of X into Y.*

Let ω denote the homotopy in $Z + C_g$ between the identity mapping of Y and the mapping g; explicitly

$$\omega_t(y) = \langle y, t\rangle, \qquad y \epsilon Y, 0 \leqq t \leqq 1.$$

If $\theta \simeq \theta'$ in $Z + C_g$ where $\theta'(X) \subset Y$ then $\theta \simeq \theta' = \omega_0\theta' \simeq \omega_1\theta' = g\theta'$. Thus a mapping f satisfying $gf \simeq \theta$ is the mapping $f = \theta'$.

If, conversely, there is an f such that $gf \simeq \theta$ then $\theta \simeq gf = \omega_1 f \simeq \omega_0 f = f$; thus f is a map of X into Y which is homotopic in $Z + C_g$ to the given map θ.

---

21 Alexandroff and Hopf, loc. cit. p. 63.

22 Ibid., p. 96-99.

23 Kuratowski, *Topologie* I, p. 127.

On choosing $Z = X$ and $\theta = 1$ in theorem 3.1 we have

THEOREM 3.3. *X is a retract of $Y + C_f$ if and only if f has a left inverse.*

On choosing $Z = X$ and $\theta = 1$ in theorem 3.2 and observing that $Z + C_g$ can be deformed into Z by the homotopy

$$\xi_t(\langle y, s\rangle) = \langle y, s + t(1 - s)\rangle, \qquad \langle x, s\rangle \in C_g,$$

$$\xi_t(z) = z, \qquad z \in Z,$$

we have

THEOREM 3.4.[24] *$Z + C_g$ can be deformed into Y if and only if g has a right inverse.*

Comparison of theorems 3.3 and 3.4 shows a curious "duality" between deformation and retraction.

(3.5) *There is a map f of X into Y such that X is a retract of $Y + C_f$ if and only if there is a map g of Y into X such that $X + C_g$ can be deformed into Y.*

This is a corollary of the more general "duality" implied by 3.1 and 3.2.

(3.6) *There is a map f of X into Y such that θ can be extended to $Y + C_f$ if and only if there is a map g of Y into Z such that θ is homotopic in $Z + C_g$ to a mapping of X into Y.*

From theorems 1.2, 2, 3.3, and 3.4 follows

THEOREM 3.7. *X is a deformation retract of $Y + C_f$ if and only if f has a 2-sided inverse.*

This theorem shows the practical equivalence of the two concepts, deformation retraction and homotopy type. To emphasize this I restate theorem 3.7.

THEOREM 3.8. *Two spaces X and Y belong to the same homotopy type if and only if they can both be imbedded in a third space W in such a way that they are both deformation retracts of W. The dimension of W need not be larger than* max $\{\dim Y, 1 + \dim X\}$.

By induction there follows

(3.9) *If the spaces $X_1, \cdots, X_k$ belong to the same homotopy type then there is a space W of which each X_i is a deformation retract. The dimension of W need not be larger than* $1 + \max_{i=1,\cdots,k} \{\dim X_i\}$, *or than* $\max_{i=1,\cdots,k} \{\dim X_i\}$ *if only one X_i has the maximum dimensionality.*

I conclude this section with an example illustrating the utility of theorem 3.7. Let T be a 2-simplex with vertices a, b, c and let E be the 2-dimensional complex resulting from T by an identification (denoted by i) of the side ab with the side bc and with the side ac. From general theorems it is known that E is contractible. I will now show how to construct, explicitly, a contraction[25] of E. Let A be a small 2-simplex in the interior of $E = i(T)$. The projection f from an interior point of A maps $X = \dot{A}$[26] onto $Y = i(\dot{T})$. Both X and Y are 1-spheres and $Y + C_f = E - (A - \dot{A})$. Since f is homotopic to a homeomorphism of X on Y, f has a 2-sided inverse. Hence, by theorem 3.7, X is a

[24] Equivalently: *Z can be deformed into $C_g - Z$ if and only if g has a right inverse.*

[25] The contractibility of this example was shown by K. Borsuk, *Über das Phänomen der Unzerlegbarkeit in der Polyedertopologie.* Comm. Math. Helv. 8 (1935), §3, p. 143.

[26] The dot denotes the boundary operation, as in Alexandroff and Hopf, loc. cit.

deformation retract of $Y + C_f$. By theorem 1.4 the retracting deformation may be chosen so that it leaves the points of X fixed throughout the deformation. Thus A is a deformation retract of E. Since A is contractible it follows that E is contractible.

4. The Hopf-Pannwitz deformations

A space A is said to be *inessential*[27] *relative to a subset* B if there is a deformation of A into a proper subset of itself such that the points of B remain fixed during the deformation.

THEOREM 4.1. *Let f be a mapping of X into Y such that $Y + C_f$ is an ANR-set and suppose that f has a right inverse. If V is a proper subset of $Y + C_f$* (which contains X) *and h is a mapping of Y into V such that $hf \simeq 1$ in V then $Y + C_f$ is inessential relative to X.*

On choosing Z of theorem 3.1 to be the set V we have that the identity map of X can be extended to a map of $Y + C_f$ into V. Let this extension be denoted by λ. Since f has a right inverse it follows from theorem 3.4 that $Y + C_f$ can be deformed into X. Hence, by 1.1, $\lambda \simeq 1$ in $Y + C_f$. Thus there is a deformation η of $Y + C_f$ such that $\eta_1 = \lambda$. By 1.3, η can be so chosen that $\eta_t \mid X = 1$ for every $0 \leqq t \leqq 1$. Thus $Y + C_f$ is inessential relative to X.

Let X be the graph consisting of two circles S_1 and S_2 joined by an arc and let Y be a 2-cell. Let f be the map of X into Y which maps the arc into a point of the boundary $\dot{Y}$[26] of Y and maps each of the circles homeomorphically onto $\dot{Y}$. Since Y is contractible, f has a right inverse. A map h satisfying the condition of theorem 4.1 is a map of Y into a point of X, where $V = X + Y + C_{f \mid (S_1 + S_2)}$. (It is easy to verify that X can be contracted in V.) Hence, by theorem 4.1, $Y + C_f$ is inessential relative to X. Let K be the space obtained from a torus by spanning a meridian with a 2-cell and an equator with a 2-cell. It is easy to see that the mapping cylinder $Y + C_f$ just constructed is a subset of K in such a way that $(K - (Y + C_f)) \cdot (Y + C_f) = X$. Hence we deduce that K is inessential. This gives a new and simple proof of a deformation theorem of Hopf and Pannwitz.[28]

If Y is a point then a mapping of Y into a point of X is a right inverse of f. Hence, on choosing $h(y) \in X$, we have

THEOREM 4.2. *If f maps the ANR-set X into a point of Y and if X can be contracted in a proper subset of $Y + C_f$ then $Y + C_f$ is inessential relative to X.*

By choosing X to be the graph described above, theorem 4.2 yields a new proof of another example of Hopf and Pannwitz.

If Y is a point and $f(X) \subset Y$ then $Y + C_f = C_f$ is called the *cone* of X. A homogeneous n-dimensional polyhedron K is said to be *closed*[29] if for some coefficient domain there is an n-cycle whose carrier[30] is K.

---

[27] Alexandroff and Hopf, loc. cit. p. 521.
[28] Ibid, p. 525.
[29] Ibid, p. 274.
[30] Ibid, p. 169 where the carrier for C is denoted by $\mid C \mid$.

THEOREM 4.3. *The homogeneous n-dimensional* ($n > 0$) *polyhedron* K *is closed if and only if* K *can not be contracted in any proper subset of its cone.*

If K cannot be so contracted that K is closed by theorem 4.2 and a theorem of Hopf and Pannwitz.[31] Suppose K were closed and could be contracted in a proper subset of the cone. Then a continuous $(n + 1)$-chain, covering a proper subset of the cone, could be found whose boundary cycle has K for its carrier. Since every n-simplex of K lies on exactly one $(n + 1)$-simplex of the cone and conversely, this is impossible.

The *absolute boundary*[32] of a homogeneous n-dimensional polyhedron K is made up of those simplexes which, for every coefficient domain, carry the boundary of every n-chain whose carrier is K.

THEOREM 4.4. *A point* p *of a homogeneous n-dimensional* ($n > 0$) *polyhedron* K *belongs to the absolute boundary of* K *if and only if the boundary of its star* $st(p)$ *can be contracted in a proper subset of* $st(p)$.[33]

Since $st(p)$ is the cone of the boundary of $st(p)$ this is a consequence[34] of theorem 4.3.

5. Special deformation retracts and homotopy type

Let

$$\rho(\langle x, s\rangle) = f(x), \qquad \langle x, s\rangle \in C_f,$$

$$\rho(y) = y, \qquad y \in Y.$$

I shall say that a retracting deformation ξ of $Y + C_f$ into X is *special* if

$$\xi_t \mid X = 1, \qquad 0 \leqq t \leqq 1 \qquad \text{and}$$

$$\rho(\xi_t(f(x))) = \rho(\xi_1(\langle x, t\rangle)).$$

A two-sided inverse g of f I will call *special* if there are homotopies F and G such that in addition to the usual conditions

$$F_0(x) = x, \qquad F_1(x) = gf(x), \qquad G_0(y) = y, \qquad G_1(y) = fg(y)$$

the condition

$$f(F_t(x)) = G_t(f(x))$$

is satisfied.

THEOREM 5. *In order that* f *have a special 2-sided inverse it is necessary and sufficient that there exist a special retracting deformation of* $Y + C_f$ *into* X.

---

[31] Ibid, p. 521.

[32] Ibid, p. 285.

[33] The star of p is the union of the closed simplexes of K which contain p. The boundary of $st(p)$ is the union of those closed simplexes of K which are contained in $st(p)$ and do not contain p.

[34] Ibid, Satz XIV, p. 285.

Suppose first that ξ is a special retracting deformation of $Y + C_f$ into X. Let

$$g(y) = \xi_1(y), \qquad y \in Y,$$
$$F_t(x) = \xi_1(\langle x, t\rangle), \qquad x \in X, \qquad 0 \leqq t \leqq 1,$$
$$G_t(y) = \rho(\xi_t(y)), \qquad y \in Y, \qquad 0 \leqq t \leqq 1.$$

Clearly $F_0(x) = \xi_1(\langle x, 0\rangle) = \xi_1(x) = x$; $F_1(x) = \xi_1(\langle x, 1\rangle) = \xi_1 f(x) = g(f(x))$; $G_0(y) = \rho(\xi_0(y)) = \rho(y) = y$; $G_1(y) = \rho(\xi_1(y)) = \rho(g(y)) = \rho(\langle g(y), 0\rangle) = f(g(y))$. Also $fF_t(x) = f(\xi_1(\langle x, t\rangle)) = \rho(\xi_1(\langle x, t\rangle)) = \rho(\xi_t(f(x))) = G_t(f(x))$.

Conversely suppose g, F and G are given satisfying $f(F_t(x)) = G_t(f(x))$. A retracting deformation ξ of C_f is defined by

$$\begin{aligned}
\xi_u(\langle x, y\rangle) &= F_{(2u+1)t}(x), && 0 \leqq t \leqq 2u/(2u+1), 0 \leqq u \leqq 1/2,\\
&= \langle F_{2u}(x), (2u+1)t - 2u\rangle, && 2u/(2u+1) \leqq t \leqq 1, 0 \leqq u \leqq 1/2,\\
&= F_{2t}(x), && 0 \leqq t \leqq 1/2, 1/2 \leqq u \leqq 1,\\
&= \langle g(f(x)), (2t-1)(2-2u)\rangle, && 1/2 \leqq t \leqq 1, 1/2 \leqq u \leqq 1.
\end{aligned}$$

It is easy to verify that this definition is consistent. When $0 \leqq u \leqq 1/2$ we have $\xi_u(f(x)) = \xi_u(\langle x, 1\rangle) = \langle F_{2u}(x), 1\rangle = f(F_{2u}(x)) = G_{2u}(f(x))$ and when $1/2 \leqq u \leqq 1$ we have $\xi_u(f(x)) = \langle gf(x), 2 - 2u\rangle$. Hence I may consistently define

$$\begin{aligned}
\xi_u(y) &= G_{2u}(y), && 0 \leqq u \leqq 1/2,\\
&= \langle g(y), 2 - 2u\rangle, && 1/2 \leqq u \leqq 1,
\end{aligned}$$

and thus extend ξ to a retracting deformation of $Y + C_f$ into X. For every $0 \leqq u \leqq 1$ we have $\xi_u(x) = \xi_u(\langle x, 0\rangle) = F_0(x) = x$. Thus $\xi_u \mid X = 1$ for every $0 \leqq u \leqq 1$. Finally I show that $\rho(\xi_t(f(x))) = \rho(\xi_1(\langle x, t\rangle))$. When $0 \leqq t \leqq 1/2$ we have $\rho(\xi_t(f(x))) = \rho(\xi_t(\langle x, 1\rangle)) = \rho(\langle F_{2t}(x), 1\rangle) = (F_{2t}(x)) = \rho(\xi_1(\langle x, t\rangle))$ and when $1/2 \leqq t \leqq 1$ we have $\rho(\xi_t(f(x))) = \rho(\xi_t(\langle x, 1\rangle)) = \rho(\langle g(f(x)), 2 - 2t\rangle) = \rho(\langle g(f(x)), 0\rangle) = \rho(\xi_1(\langle x, t\rangle))$.

6. Mappings with ε-inverses for every ε

A homotopy ξ in a metric space B is called an ε-homotopy[35] if $d(\xi_{t_1}(x), \xi_{t_2}(x)) < \varepsilon$ for every t_1, $t_2 \in [0, 1]$ and $x \in \xi^{-1}(B)$. Accordingly if X and Y are metric spaces and f and g are mappings, of X into Y and Y into X respectively, such that $gf \simeq_\varepsilon 1$ ($\simeq_\varepsilon$ denotes ε-homotopy) then I will call g a *left ε-inverse* of f and f a *right ε-inverse* of g. The property of having a left or right ε-inverse for every $\varepsilon > 0$ is topological.

Assume now that the mapping cylinder $Y + C_f$ is metrizable and has been metrized. This is the case, for instance, if X and Y are compacta.

THEOREM 6.1. *A mapping f of X into Y has a left ε-inverse for every $\varepsilon > 0$ if and only if f is a homeomorphism and for every $\varepsilon > 0$ the identity mapping of $f(X)$ is ε-homotopic in $f(X)$ to a map extendable to Y.*

---

[35] Ibid, p. 343.

If g^{ε} is a left ε-inverse of f then $f(x_1) = f(x_2)$ implies that $d(x_1, x_2) \leqq d(x_1, g^{\varepsilon}(f(x_1))) + d(g^{\varepsilon}(f(x_2)), x_2) < 2\varepsilon$. Hence if f has left ε-inverses for every $\varepsilon > 0$ then f must be one to one. If $\{x_k\} \subset X$ and $f(x_k) \to f(x_0)$ then $g^{\varepsilon}(f(x_k)) \to g^{\varepsilon}(f(x_0))$ for every $\varepsilon > 0$ as $k \to \infty$ and $g^{\varepsilon}(f(x_k)) \to x_k$ uniformly in k as $\varepsilon \to 0$. Hence $x_k \to x_0$ so that f is a homeomorphism. For every $\varepsilon > 0$ there is a $\delta > 0$ such that $fg^{\delta}ff^{-1} \simeq_{\varepsilon} f1f^{-1} = 1$ in $f(X)$. Hence the mapping $fg^{\delta} \mid f(X)$ is extendable to Y and also ε-homotopic to the identity.

It is no loss of generality, in the proof of the converse, to assume that $X \subset Y$ and $f = 1$. The maps g^{ε} of Y into X such that $g^{\varepsilon} \mid X \simeq_{\varepsilon} 1$ in X are easily seen to be left ε-inverses of f.

(6.2) *If X is, in addition, an ANR-set then f has a left ε-inverse if and only if f is a homeomorphism and $f(X)$ is a retract of Y.*

This follows from (6.1) and a theorem[14] of Borsuk-Kuratowski-Dowker.

THEOREM 6.3. *In order that f have a right ε-inverse for every $\varepsilon > 0$ it is necessary and sufficient that Y be ε-deformable[36] into $C_f - Y$ for every $\varepsilon > 0$.*

If, for every $\delta > 0$, g^{δ} is a right δ-inverse and G^{δ} is a δ-homotopy such that $G_0^{\delta}(y) = y$ and $G_1^{\delta}(y) = f(g^{\delta}(y))$ then, for preassigned ε and sufficiently small δ, the deformation ξ^{δ} defined by

$$\begin{aligned} \xi_t^{\delta}(y) &= G_{2t}^{\delta}(y), \qquad y \in Y, \qquad 0 \leqq t \leqq 1/2, \\ &= \langle g^{\delta}(y), 1 - (2t - 1)\delta \rangle, \qquad y \in Y, \qquad 1/2 \leqq t \leqq 1, \end{aligned}$$

is an ε-deformation of Y into $C_f - Y$.

Let $\nu(\langle x, t \rangle) = x$ for every $\langle x, t \rangle \in C_f - Y$. If, for every $\delta > 0$, ξ^{δ} is a δ-deformation of Y into $C_f - Y$ then, for preassigned ε and sufficiently small δ, the mapping g of Y into X defined by

$$g(y) = \nu(\xi_1^{\delta}(y)), \qquad y \in Y,$$

is a right ε-inverse of f.

Note that the condition of theorem 6.3 implies that $\overline{f(X)} = Y$. Hence

(6.4) *If X is compact then f has a 2-sided ε-inverse for every $\varepsilon > 0$ if and only if f is a homeomorphism of X onto Y.*

7. Analysis of n-homotopy type

Mappings f and g of A into B are called *n-homotopic*[37] if for every n-dimensional polyhedron P and mapping ϕ of P into A the mappings $f\phi$ and $g\phi$ are homotopic. The symbol $\simeq^n$ will denote n-homotopy.

THEOREM 7.1. *Let X, Y and Z be topological spaces and let θ be a mapping of X into Z. If f is a mapping of X into Y then there is a mapping g of Y into Z satisfying $gf \simeq^n \theta$ if and only if θ is n-homotopic to a mapping which can be extended to $Y + C_f$.*

If θ^* is a mapping of $Y + C_f$ into Z such that $\theta^* \mid X \simeq^n \theta$ then the mapping $g = \theta^* \mid Y$ satisfies $gf \simeq^n \theta$. In fact $gf = \theta^* \mid \langle X, 1 \rangle \simeq \theta^* \mid \langle X, 0 \rangle = \theta^* \mid X \simeq^n \theta$.

[36] I.e. the deformation is an ϵ-homotopy.

[37] R. H. Fox, *On the Lusternik-Schnirelmann Category*, Annals of Math. 42 (1941), p. 344.

Suppose, conversely, that $gf \simeq^n \theta$. Define

$$\theta^*(\langle x, t\rangle) = g(f(x)), \qquad \langle x, t\rangle \in C_f,$$
$$\theta^*(y) = g(y), \qquad y \in Y.$$

Then $\theta^* X = gf \simeq^n \theta$.

THEOREM 7.2. *Let X, Y, Z and θ be as in theorem 7.1. If g is a mapping of Y into Z then there is a mapping f of X into Y satisfying $gf \simeq^n \theta$ if and only if θ is n-homotopic in $Z + C_g$ to a mapping of X into Y.*

The proof can be constructed from that of theorem 3.2 by changing the appropriate homotopies to n-homotopies.

If $gf \simeq^n 1$ the mapping g will be called a left n-homotopy inverse and f will be called a right n-homotopy inverse. If f has a 2-sided n-homotopy inverse X and Y will be said to belong to the same n-homotopy type. If X and Y are compact ANR-sets then g is a left (right) inverse of f if and only if g is a left (right) n-homotopy inverse of f for every $n < 1 + \dim X$ (for every $n < 1 + \dim Y$).[38]

(7.3) *f has a left n-homotopy inverse if and only if the identity mapping of X is n-homotopic to a map which is extendable to $Y + C_f$.*

(7.4) *g has a right n-homotopy inverse if and only if $Z + C_g$ can be n-deformed into Y.*

By the argument of (1.1) it follows from (7.3) and (7.4) that

(7.5) *X and Y belong to the same n-homotopy type if and only if there is a space $W \supset X + Y$, of which Y is a deformation retract, and a mapping h of W into X such that $h \simeq^n 1$ in W and $h \mid X \simeq^n 1$ in X.*

For example a 2-sphere X and a point Y belong to the same 1-homotopy type. Here a space W is a 2-cell of which X is the boundary and Y is the center. Spaces which belong to the same n-homotopy type have isomorphic k-dimensional homotopy groups for $k \leqq n$.

UNIVERSITY OF ILLINOIS.

---

[38] Ibid, theorem 13, p. 344.

ANNALS OF MATHEMATICS
Vol. 44, No. 1, January, 1943

ON THE DEFORMATION RETRACTION OF SOME FUNCTION SPACES ASSOCIATED WITH THE RELATIVE HOMOTOPY GROUPS

BY RALPH H. FOX

(Received October 1, 1942)

The homotopy group $\pi^{n+1}(B, C) = \pi^{n+1}(B, C, d)$, $(n \geqq 1)$, of an arcwise connected[1] space B relative[2] to an arcwise connected[1] subset C with base point $d \in C$ may be defined as the fundamental group of a certain space[3] $\mathfrak{F}^n(B, C, d)$. The group $\pi^{n+1}(B, C)$ is independent of the base point d in the sense that $\pi^{n+1}(B, C, d_1)$ and $\pi^{n+1}(B, C, d_2)$ are isomorphic. Hurewicz demonstrates this by showing that $\mathfrak{F}^n(B, C, d_1)$ and $\mathfrak{F}^n(B, C, d_2)$ belong to the same homotopy type. According to my generalization[4] of Whitehead's theorem these function spaces are therefore deformation retracts of some containing space W. The containing space W constructed by this method is a subset of the function space $\mathfrak{F}^n(B, C, D)$, where D is the arc, with end-points d_1 and d_2 which appears in Hurewicz' (unpublished) proof. This suggests that $\mathfrak{F}^n(B, C, d_1)$ and $\mathfrak{F}^n(B, C, d_2)$ may be deformation retracts of $\mathfrak{F}^n(B, C, D)$ itself.

It will be shown below that this is indeed the case, at least if B and C are compact ANR-sets. (The reader will note that the reasoning can now be reversed to deduce the independence of $\pi^{n+1}(B, C)$ of the base point.) Furthermore deformation retraction of B or C induces deformation retraction of $\mathfrak{F}^n(B, C, d)$, hence leaves $\pi^{n+1}(B, C)$ unaltered. Before plunging into the proof I generalize the spaces $\mathfrak{F}^n(B, C, D)$ by (a) generalizing the antecedent cells and cell-boundaries to arbitrary topological spaces, and (b) removing the dependence of $\mathfrak{F}^n$ on the obviously irrelevant number three. Thus we might loosely describe the investigation as a study of the relationship between deformation retraction in the image space and deformation retraction of the function space.

---

[1] This slight restriction, which is not required for the definition of the group, simplifies our discussion.

[2] The absolute homotopy group $\pi^{n+1}(B) = \pi^{n+1}(B, d)$, whose definition may be obtained from that of the relative group by identifying B with d, may be discussed analogously.

[3] The set of continuous functions defined on a topological space X with values in a metric space Y is denoted, as usual, by the symbol Y^X. If X is compact or Y is bounded the well known formula

$$d(f, g) = \sup_{x \in A}\{d(f(x), g(x))\}$$

makes Y^X a metric space. If either X or Y is compact then topologically equivalent metrics of Y induce equivalent metrics of Y^X so that the topology of Y^X is then independent of the metrization of the topological space Y.

Let E^n denote the n-cell: $0 \leqq x_i \leqq 1$, $i = 1, 2, \cdots, n$ and $\dot{E}^n$ its boundary: $\prod_{i=1}^n x_i(1 - x_i) = 0$. If D is any subset of C, the symbol $\mathfrak{F}^n(B, C, D)$ denotes the subset of B^{E^n} which consists of those mappings f for which $f(\dot{E}^n) \subset C$ and $f(\dot{E}^n - E^{n-1}) \subset D$.

[4] R. H. Fox, *On Homotopy Type and Deformation Retraction*, this volume, p. 45, theorem 3.8.

The three lemmas are contributions to the theory of fibre spaces[5] and, except for notation, are independent of the rest of the paper. With reference to lemma 3 it should be pointed out that a certain theorem of Borsuk[6], published four years before fibre spaces[5] had been discovered, is, when restated, seen to be a far-reaching result on fibre spaces. Borsuk's fibre theorem reads: If B is a compact ANR-set and A' is closed in A then the operation $f \to f \mid A'$[7], $f \in B^A$ is a fibre mapping of B^A into $B^{A'}$. I have recently discovered a very simple proof of this theorem which will appear elsewhere together with a discussion of its place in fibre space theory.

Let A_λ (λ a non-negative integer) denote a topological space and B_λ a metric space with either A_λ compact or B_λ bounded. Consider a decreasing sequence

$$A_\lambda \supset A_{\lambda-1} \supset \cdots \supset A_0 \supset \cdots \supset A_{-\mu}, \qquad (0 \leqq \mu \leqq \infty)$$

of subsets of A_p and a corresponding sequence

$$B_\lambda \supset B_{\lambda-1} \supset \cdots \supset B_0 \supset \cdots \supset B_{-\mu}$$

of subsets of B_λ. Let B_0' be a subset of B_0 which contains B_{-1} if $\mu \neq 0$. For $0 \leqq i \leqq \lambda$ I shall use the symbol $\mathfrak{F}_i$ to denote the subset of $B_i^{A_i}$ which consists of those mappings f for which

$$f(A_j) \subset B_j \quad \text{for} \quad -\mu \leqq j \leqq i,$$

and the symbol $\mathfrak{F}_i'$ for the subset of $\mathfrak{F}_i$ which consists of those mappings $f \in \mathfrak{F}_i$ which satisfy the additional requirement

$$f(A_0) \subset B_0'.$$

Let h be a deformation of B_0 into B_0' which deforms each of the subsets $B_{-1}, \cdots, B_{-\mu}$ within itself. Thus $h \in B_0^{B_0 \times [0,1]}$ such that

$$\text{(1)} \qquad \begin{cases} h_0(y) &= y \quad \text{for} \quad y \in B_0, \\ h_1(B_0) &\subset B_0', \\ h_t(B_{-\nu}) &\subset B_{-\nu} \quad \text{for} \quad 0 \leqq t \leqq 1, \quad -\mu \leqq -\nu \leqq 0. \end{cases}$$

The existence of this deformation is assumed from now on.

THEOREM 1. *If B_0 is closed and $B_1, \cdots, B_\lambda$ are ANR-sets[8] closed in B_λ then $\mathfrak{F}_\lambda$ can be deformed into $\mathfrak{F}_\lambda'$.*

---

[5] W. Hurewicz and N. E. Steenrod, *Homotopy Relations in Fibre Spaces*, Proc. Nat. Acad. 27 (1941), 61–64. The earlier definitions of H. Whitney (sphere-spaces, sphere-bundles and fibre bundles) required the fibres to belong to the same topological type. Other definitions—see B. Eckmann, *Zur Homotopietheorie gefaserter Räume*, Comm. Math. Helv. 14 (1941), 141–192, for references—require compactness assumptions.

[6] K. Borsuk, *Sur les prolongements des transformations continues*, Fund. Math. 28 (1937), 99–110.

[7] R. H. Fox, *On Homotopy Type and Deformation Retraction*. loc. cit., footnote 10.

[8] R. H. Fox, *A Characterization of Absolute Neighborhood Retracts*, Bull. Am. Math. Soc. 48 (1942), 271–275.

Since B_0 is closed in the ANR-set B_1 the deformation h of B_0 can be extended to a deformation of B_1 (in itself)[9]. Since B_i is closed in the ANR-set B_{i+1} ($i = 1, 2, \cdots, \lambda - 1$) the same argument shows that when h has been extended stepwise to a deformation of B_i it can be further extended to a deformation of B_{i+1}. Let h^* denote the deformation of B_λ which is the final result of this sequence of extensions. Thus $h^* \in B_\lambda^{B_\lambda \times [0,1]}$ such that

$$(2) \qquad \begin{cases} h_0^*(y) = y & \text{for} \quad y \in B_\lambda \\ h_1^*(B_0) \subset B_0' \\ h_t^*(B_i) \subset B_i & \text{for} \quad 0 \leqq t \leqq 1, \quad -\mu \leqq i \leqq \lambda. \end{cases}$$

For every $x \in A_0$ and $f \in \mathfrak{F}_\lambda$ define

$$(3) \qquad \phi_t^f(x) = h_t^*(f(x)).$$

The function ϕ is a deformation of $\mathfrak{F}_\lambda$ into $\mathfrak{F}_\lambda'$.

THEOREM 2. *If $A_0, \cdots, A_\lambda$ are ANR-sets closed in A_λ then $\mathfrak{F}_\lambda$ is deformable into $\mathfrak{F}_\lambda'$.*

First I construct a deformation r of A_λ which is a neighborhood retracting deformation of A_0 and which deforms each $A_i (-\mu \leqq i \leqq \lambda)$ within itself. The construction is inductive; the mapping

$$^0R_t(x) = x \quad \text{for} \quad x \in A_0$$

is a deformation of A_0 which is a neighborhood retracting deformation of A_0 and which deforms each $A_i(-\mu \leqq i \leqq 0)$ within itself. Since A_0 is a neighborhood retract of A_λ there is a closed neighborhood U_0 of A_0 in A_λ and a retraction ρ of U_0 onto A_0. Suppose that $^jR \in A_j^{A_j \times [0,1]}$ ($0 \leqq j < \lambda$) such that

$$\begin{aligned} {}^jR_0(x) &= x \quad \text{for} \quad x \in A_j, \\ {}^jR_t(A_i) &\subset A_i \quad \text{for} \quad 0 \leqq t \leqq 1 \quad \text{and} \quad -\mu \leqq i \leqq j, \\ {}^jR_1(x) &= \rho(x) \quad \text{for} \quad x \in U_j \cdot A_j, \end{aligned}$$

where U_j is a closed neighborhood of A_0 in U_0 (hence in A_λ). Define

$$\begin{aligned} {}^jS_0(x) &= x \quad \text{for} \quad x \in A_{j+1}, \\ {}^jS_t(x) &= {}^jR_t(x) \quad \text{for} \quad (x, t) \in A_j \times [0, 1], \\ {}^jS_1(x) &= \rho(x) \quad \text{for} \quad x \in U_j \cdot A_{j+1}. \end{aligned}$$

Since $A_{j+1} \times [0] + A_j \times [0, 1] + U_j \cdot A_{j+1} \times [1]$ is closed in $A_{j+1} \times [0, 1]$ and A_{j+1} is an ANR-set, ^jS can be[10] extended to a map $^jS^*$ of a neighborhood V_j (into A_{j+1}). Let U_{j+1} be a closed neighborhood of A_0 in U_j (hence in A_λ) such

[9] R. H. Fox, *On Homotopy Type and Deformation Retraction.* loc. cit. footnote 14.

[10] R. H. Fox, *A Characterization of Absolute Neighborhood Retracts,* loc. cit. p. 273.

that

$$U_{j+1} \cdot A_{j+1} \times [0, 1] \subset V_j,$$

and define

$$^jT_0(x) = x \quad \text{for} \quad x \in A_{j+1},$$
$$^jT_t(x) = {}^jS_t^*(x) \quad \text{for} \quad (x, t) \in U_{j+1} \cdot A_{j+1} \times [0, 1].$$

Since ^jT is homotopic to the identity mapping of the closed set $A_{j+1} \times [0] + U_{j+1} \cdot A_{j+1} \times [0, 1]$ and $A_{j+1} \times [0, 1]$ is an ANR it follows[9] that ^jT can be extended to $A_{j+1} \times [0, 1]$. Let ^{j+1}R denote the extended mapping. But $^{j+1}R \in A_{j+1}^{A_{j+1} \times [0,1]}$ such that

$$^{j+1}R_0(x) = x \quad \text{for} \quad x \in A_{j+1},$$
$$^{j+1}R_t(A_i) \subset A_i \quad \text{for} \quad 0 \leqq t \leqq 1 \quad \text{and} \quad -\mu \leqq i \leqq j + 1,$$
$$^{j+1}R_1(x) = \rho(x) \quad \text{for} \quad x \in U_{j+1} \cdot A_{j+1}.$$

This completes the induction; let $r = {}^\lambda R$ so that

$$(4) \qquad \begin{cases} r_0(x) = x & \text{for} \quad x \in A_\lambda, \\ r_t(A_i) \subset A_i & \text{for} \quad 0 \leqq t \leqq 1 \quad \text{and} \quad -\mu \leqq i \leqq \lambda, \\ r_1(x) = \rho(x) & \text{for} \quad x \in U_\lambda. \end{cases}$$

Let $\delta \in [0, 1]^{A_\lambda}$ such that $\delta(\overline{A_\lambda - U_\lambda}) = 0$ and $\delta(A_0) = 1$[11]. A deformation of $\mathfrak{F}_\lambda$ into $\mathfrak{F}'_\lambda$ is defined by the formulae

$$(5) \qquad \begin{aligned} \phi_t^f(x) &= f(r_{2t}(x)) \quad \text{for} \quad 0 \leqq t \leqq 1/2, x \in A_\lambda, \\ &= h_{(2t-1)\delta(x)}(f(r_1(x))) \quad \text{for} \quad 1/2 \leqq t \leqq 1, x \in U_\lambda, \\ &= f(r_1(x)) \quad \text{for} \quad 1/2 \leqq t \leqq 1, x \in A_\lambda - U_\lambda, \end{aligned}$$

where $f \in \mathfrak{F}_\lambda$.

Given a topological space X and metric space Y I shall say that a mapping $\pi \in Y^X$ is a *fibre mapping* if X is a fibre space[5] over $\pi(X) \subset Y$ relative to π. This implies, of course, that $\pi(X)$ is open and closed in Y.[5]

LEMMA 1. *If π is a fibre mapping of X into Y and X' is the complete inverse image of some subset Y' of Y then $\pi \mid X'$[7] is a fibre mapping of X' into Y.*

The proof is immediate.

LEMMA 2. *If π_1 is a fibre mapping of X into Y and π_2 is a fibre mapping of Y into Z whose slicing function is uniformly continuous in y and z together then $\pi_2\pi_1 \in Z^X$ is a fibre mapping.*

According to the definition there exist $\epsilon_1, \epsilon_2 > 0$ and slicing functions ϕ_1, ϕ_2 such that

$$\phi_1(x, y) \in \pi_1^{-1}(y) \text{ and is defined whenever } d(\pi_1(x), y) < \epsilon_1,$$

$$\phi_2(y, z) \in \pi_2^{-1}(z) \text{ and is defined whenever } d(\pi_2(y), z) < \epsilon_2,$$

$$\phi_1(x, \pi_1(x)) = x \quad \text{and} \quad \phi_2(y, \pi_2(y)) = y.$$

---

[11] Urysohn's Lemma.

Choose $\epsilon < \epsilon_2$ so small that $d(\phi_2(x, z), \phi_2(x, z')) < \epsilon_1$ whenever $d(z, z') < \epsilon$; this is possible because of the uniform continuity of ϕ_2. Let

$$\phi(x, z) = \phi_1(x, \phi_2(\pi_1(x), z)). \tag{6}$$

This is defined whenever $d(\pi_2\pi_1(x), z) < \epsilon$. Furthermore

$$\begin{aligned} \pi_2\pi_1\phi(x, z) &= \pi_2(\pi_1\phi_1(x, \phi_2(\pi_1(x), z))) \\ &= \pi_2(\phi_2(\pi_1(x), z)) \\ &= z, \end{aligned}$$

and

$$\begin{aligned} \phi(x, \pi_2\pi_1(x)) &= \phi_1(x, \phi_2(\pi_1(x), \pi_2\pi_1(x))) \\ &= \phi_1(x, \pi_1(x)) \\ &= x. \end{aligned}$$

Let π_{ij} be the mapping of $\mathfrak{F}_i$ into $\mathfrak{F}_j (i \geqq j)$ defined by $\pi_{ij}(f_i) = f_i \mid A_j{}^7, f_i \in \mathfrak{F}_i$.

LEMMA 3. *If $A_0, \cdots, A_{\lambda-1}$ are closed and $B_1, \cdots, B_\lambda$ are compact ANR-sets then $\pi_{\lambda 0}$ is a fibre mapping*[12].

By the proof of a theorem of Borsuk[6] [in particular, formulae (8) p. 101 (which should read: $\varphi^*(p) = 1 - r_{01}[\bar{\varphi}(p) + \overset{*}{\varphi_0}(p) - \bar{\varphi}_0(p)]$), l. 11 p. 102 (which should read $\bar{\varphi}(x) = \{\varphi^{(*,n)}(x)\}$) and l. 6. p. 103] the mapping $f_i \to f_i \mid A_{i-1}{}^7, f_i \in B_i^{A_i}$, $1 \leqq i \leqq \lambda$ is a fibre mapping. Moreover [the previously mentioned formulae, (1) p. 100 and modification of the proof of theorem 2 to the extent of replacing the neighborhood U last line p. 102 by a closed neighborhood so that the retraction r is uniformly continuous] the slicing function for this fibre mapping is uniformly continuous.

But the inverse image of $\mathfrak{F}_{i-1}$ under the mapping is precisely $\mathfrak{F}_i$. Hence, by lemma 1, $\pi_{i\,i-1}$ is a fibre mapping. Hence, by lemma 2, $\pi_{\lambda 0} = \pi_{10}\pi_{21} \cdots \pi_{\lambda\lambda-1}$ is a fibre mapping.

THEOREM 3. *If $A_0, \cdots, A_{\lambda-1}$ and B_0 are closed and $B_1, \cdots, B_\lambda$ are compact ANR-sets and if $h_t(y) = y$ for every $y \in B_0'$, $0 \leqq t \leqq 1$ then $\mathfrak{F}_\lambda'$ is a deformation retract of $\mathfrak{F}_\lambda$.*

Let $\psi \in \mathfrak{F}_0^{\mathfrak{F}_\lambda \times [0,1]}$ defined by

$$\psi_t^f(x) = h_t f(x) \quad \text{for} \quad 0 \leqq t \leqq 1, \quad x \in A_0 \quad \text{and} \quad f \in \mathfrak{F}_\lambda. \tag{7}$$

I shall show that $\psi(\mathfrak{F}_\lambda \times [0, 1]) \subset \pi_{\lambda 0}(\mathfrak{F}_\lambda)$. In fact

$$\begin{aligned} \psi_t^f &= \pi_{00}(\psi_t^f), \\ \psi_0^f(x) &= f(x) \quad \text{for} \quad x \in A_0, \end{aligned}$$

---

[12] Since the image set of a fibre mapping is open and closed it follows from R. H. Fox, *A Characterization of Absolute Neighborhood Retracts*, loc. cit. footnote 3 that the compactness of B_λ is essential to this lemma.

and $\psi_t^f \in \mathfrak{F}_0$ for every fixed $t \in [0, 1]$. Suppose, inductively, that for every fixed $f \in \mathfrak{F}_\lambda$ there is a map ${}^i\psi = {}^i\psi^f \in B_i^{A_i \times [0,1]}$, $0 \leqq i < \lambda$, such that

$$\pi_{i0}({}^i\psi_t^f) = {}^0\psi_t^f = \psi_t^f \qquad \text{for } 0 \leqq t \leqq 1,$$

$${}^i\psi_0^f(x) = f(x) \qquad \text{for } x \in A_i,$$

$${}^i\psi_t^f \in \mathfrak{F}_i \qquad \text{for every fixed } t \in [0, 1].$$

Let

$${}^i\xi_t^f(x) = \begin{cases} {}^i\psi_t^f(x) & \text{for } (x, t) \in A_i \times [0, 1], \\ f(x) & \text{for } (x, t) \in A_{i+1} \times [0]. \end{cases}$$

Since ${}^i\xi^f$ is homotopic to the map ${}^i\eta^f$ defined by

$${}^i\eta_t^f(x) = f(x) \quad \text{for} \quad (x, t) \in A_{i+1} \times [0] + A_i \times [0, 1],$$

since ${}^i\eta^f$ can be extended to $A_{i+1} \times [0, 1]$ and since ${}^i\xi^f$ maps the closed subset $A_{i+1} \times [0] + A_i \times [0, 1]$ of $A_{i+1} \times [0, 1]$ into the compact ANR-set B_{i+1}, it follows[9] that ${}^i\xi^f$ can be extended to a map ${}^{i+1}\psi^f$ of $A_{i+1} \times [0, 1]$ into B_{i+1}. But $\pi_{i+1\,0}({}^{i+1}\psi_t^f) = \pi_{i0}(\pi_{i+1\,i}({}^{i+1}\psi_t^f)) = \psi_t^f$ for $0 \leqq t \leqq 1$,

$${}^{i+1}\psi_0^f(x) = {}^i\xi_0^f(x) = f(x) \quad \text{for} \quad x \in A_{i+1},$$

$${}^{i+1}\psi_t^f \in \mathfrak{F}_{i+1} \quad \text{for every} \quad t \in [0, 1],$$

and this completes the induction.

Since B_0 is compact, $\psi \in \mathfrak{F}_0^{\mathfrak{F}_\lambda \times [0,1]}$ is a uniform homotopy[5]. By the previous lemma $\pi_{\lambda 0}$ is a fibre mapping. Therefore, since $\psi(\mathfrak{F}_\lambda \times [0, 1]) \subset \pi_{\lambda 0}(\mathfrak{F}_\lambda)$, it follows by the covering homotopy theorem[5] that there is a mapping $\phi \in \mathfrak{F}_\lambda^{\mathfrak{F}_\lambda \times [0,1]}$ such that

$$\phi_0^f = f \quad \text{for every} \quad f \in \mathfrak{F}_\lambda,$$

$$\pi_{\lambda 0}(\phi) = \psi,$$

$$\phi_1^f \in \mathfrak{F}_\lambda' \quad \text{for every} \quad f \in \mathfrak{F}_\lambda.$$

Thus ϕ is a deformation of $\mathfrak{F}_\lambda$ (within itself) into $\mathfrak{F}_\lambda'$.

If $f \in \mathfrak{F}_\lambda'$ then $f(x) \in B_0'$ for every $x \in A_0$, hence $\psi_t^f = h_t f$ is independent of t. Hence, by the covering homotopy theorem[5], ϕ_t^f is independent of t. Hence, for every $f \in \mathfrak{F}_\lambda'$,

$$\phi_t^f = \phi_0^f = f.$$

Hence ϕ is a retracting deformation (which leaves the points of $\mathfrak{F}_\lambda'$ invariant).

COROLLARY. *$\mathfrak{F}_\lambda'$ is a deformation retract of $\mathfrak{F}_\lambda$ if $A_0, \cdots, A_{\lambda-1}$ are closed, B_0, B_0', $B_1, \cdots, B_\lambda$ are compact ANR-sets and h is a retracting deformation.*

For under these conditions h may be assumed to satisfy[13] the condition of the preceding theorem in addition to (1).

UNIVERSITY OF ILLINOIS, URBANA, ILL.

[13] Ibid, Theorem 1.4.

Annals of Mathematics
Vol. 44, No. 1, January, 1943

ON PERMUTATION GROUPS OF PRIME DEGREE AND RELATED CLASSES OF GROUPS

By Richard Brauer*

(Received June 17, 1942)

Introduction

The transitive permutation groups of prime degree p appear as the Galois groups of the irreducible algebraic equations $f(x) = 0$ of degree p. This is the reason that these groups have been the subject of a large number of investigations.[1] However, only few results of a general nature have been obtained. In the present paper, the theory of group representations[2] will be applied in order to derive some new theorems concerning the structure of these groups. Actually, the method can be used for the study of a wider class of groups, viz. the groups $\mathfrak{G}$ of finite order g which have the following property:

(*) *The group* $\mathfrak{G}$ *contains elements* P *of prime order* p *which commute only with their own powers* P^i.

It is clear that transitive permutation groups of degree p have the property (*). Secondly, the doubly transitive permutation groups of degree $p - 1$ are of this type.[3] A third example is furnished by the irreducible linear groups in a p-dimensional vector space whose center consists of the unit element only, in particular by the simple linear irreducible groups in p dimensions (cf. section 7).

It is easily seen (section 1) that the order g of a group $\mathfrak{G}$ with the property (*) is of the form

$$(1) \qquad g = (p - 1)p(1 + np)/t$$

where t and n are integers and where t divides $p - 1$. The group $\mathfrak{G}$ contains exactly $1 + np$ conjugate subgroups of order p, and each of them has a normalizer of order $p(p - 1)/t$. In section 2, the normal subgroups of $\mathfrak{G}$ are studied, in particular the first commutator-subgroup $\mathfrak{G}'$ and the second commutator-subgroup $\mathfrak{G}''$ of $\mathfrak{G}$. Two cases must be distinguished:

Case I. *The group* $\mathfrak{G}$ *contains a normal subgroup* $\mathfrak{S}$ *of order* $1 + np$.

We shall show that $\mathfrak{G}/\mathfrak{S}$ then is a metacyclic group of order $p(p - 1)/t$; the group $\mathfrak{S}$ possesses an outer automorphism of order p which leaves only the

* Fellow of the John Simon Guggenheim Memorial Foundation.

[1] We may mention here the work of Mathieu, C. Jordan, Sylow, Frobenius, Burnside, G. A. Miller. Cf. also E. Pascal, Repertorium der höheren Mathematik, Vol. I, part 1, 2nd German edition, Leipzig 1910.

[2] In this paper, the notation "representation of a group" always means a representation of the group by linear transformations of a vector space over the field of complex numbers (or an algebraically closed field of characteristic 0). By a "vector-space" we always mean a vector-space over this field.

[3] For this class of groups, cf. G. Frobenius, Sitzungsberichte der Preussischen Akademie, Berlin 1902, p. 351.

unit element fixed. For $t < p - 1$, we have $\mathfrak{S} = \mathfrak{G}''$, and $\mathfrak{G}'$ has the order $p(1 + np)$. For $t = p - 1$, we have $\mathfrak{S} = \mathfrak{G}'$. Unless n is of the form

$$n = u + m + ump \qquad (u, m \text{ positive integers}), \tag{2}$$

$\mathfrak{S}$ is a minimal normal subgroup of $\mathfrak{G}$ (for $n \neq 0$).

This case I is of relatively small interest. In particular, when $\mathfrak{G}$ is a transitive permutation group of degree p, $\mathfrak{S}$ consists in this case I only of the unit element 1. If $\mathfrak{G}$ is a doubly transitive group of degree $p + 1$ or an irreducible linear group in a p-dimensional space with center 1, then $\mathfrak{S}$ must be abelian.

Case II. *The group $\mathfrak{G}$ does not contain a normal subgroup of order $1 + np$.*

Here, we shall have $\mathfrak{G}' = \mathfrak{G}''$. The group $\mathfrak{G}'$ itself satisfies the condition (*); its order g' is of the form

$$g' = (p - 1)p(1 + np)/t' \tag{3}$$

where n is the same number as in (1). The number t' divides $p - 1$ and is divisible by t; we have $t' \neq p - 1$. If n is not of the form (2), in particular, if $n < p + 2$, then $\mathfrak{G}'$ is simple.

In the later sections, we shall assume that $\mathfrak{G}$, besides condition (*), satisfies the following condition

(**) *The commutator-subgroup $\mathfrak{G}'$ of $\mathfrak{G}$ is equal to $\mathfrak{G}$.*

By this condition (**), groups $\mathfrak{G}$ for which we have case I are excluded. If we have case II, the group $\mathfrak{G}'$ satisfies both conditions (*) and (**), and our theory can be applied to $\mathfrak{G}'$. From $\mathfrak{G}'$, the group $\mathfrak{G}$ can be obtained by a cyclic extension; the value of n remains unchanged.

Our main result (section 5), is: If a group $\mathfrak{G}$ satisfies the conditions (*) and (**), and if $n \geqq (p + 3)/2$, then n can be represented by the following rational function $F(p, u, h)$

$$n = F(p, u, h) = \frac{puh + u^2 + u + h}{u + 1} \tag{4}$$

where u and h are positive integers, and where $u + 1$ divides $h(p - 1)$. If $\mathfrak{G}$ satisfies the conditions (*) and (**), and if $n < (p + 3)/2$, we must have one of the following two cases:

(a) $n = 1, \quad t = 2, \quad \mathfrak{G} = LF(2, p), \qquad (p > 3).$

(b) $n = (p - 3)/2, \quad t = (p - 1)/2, \quad \mathfrak{G} = LF(2, 2^\mu)$ where $p = 2^\mu + 1$

is a Fermat prime, $p > 3$.[4]

In a later paper, the values n with $(p + 3)/2 \leqq n \leqq p + 2$ will be discussed.

It had been shown by Frobenius that $LF(2, p)$ is the only simple group of order $p(p - 1)(p + 1)/2$. In section 6 we drop the assumption (*) and prove that the groups $\mathfrak{G} = LF(2, p)$ and $LF(2, 2^\mu)$. $(2^\mu + 1 = p)$ are the only simple

[4] That permutation groups of degree p with the value $n = (p - 3)/2$ exist for these primes p, was mentioned by Frobenius, loc. cit.

groups of an order $p(p - 1)(1 + mp)/\tau$ with $m < (p + 3)/2$, (p a prime, τ, m not-negative integers, $\tau \mid (p - 1)$); if for a simple group of this order we have $m \geqq (p + 3)/2$, then m must be of the form $m = F(p, u, h)$ where u and h are positive integers.

1. Preliminary remarks

Let $\mathfrak{G}$ be a group of finite order g which satisfies the condition (*), i.e. which contains elements P of prime order p whose centralizer consists of the powers of P only. If $\mathfrak{P}$ is a p-Sylow subgroup of $\mathfrak{G}$ which contains P, then the order of $\mathfrak{P}$ cannot be larger than p, since otherwise the order of the centralizer of P in $\mathfrak{P}$ would be larger than p. Hence $g \not\equiv 0 \pmod{p^2}$, $\mathfrak{P} = \{P\}$. The number of subgroups conjugate to $\mathfrak{P}$ is of the form $1 + np$ where n is a non-negative integer. The order of the normalizer $\mathfrak{N} = \mathfrak{N}(\mathfrak{P})$ of $\mathfrak{P}$ then is $g/(1 + np)$. But since $\mathfrak{P}$ is a cyclic group of order p, and since $\mathfrak{N}$ also satisfies the condition (*), we readily see that $\mathfrak{N}$ can be generated by $\mathfrak{P}$ and another element Q such that

$$P^p = 1, \qquad Q^q = 1, \qquad Q^{-1}PQ = P^{\gamma^t} \tag{5}$$

where γ is a primitive root (mod p), and where t and q are positive integers such that

$$tq = p - 1. \tag{6}$$

The group $\mathfrak{G}$ then contains exactly t classes of conjugate elements of order p. For the order of $\mathfrak{G}$, we obtain

$$g = (p - 1)p(1 + np)/t = qp(1 + np). \tag{7}$$

Hence we have

THEOREM 1. *If $\mathfrak{G}$ is a group of finite order g which contains an element P of prime order p which commutes only with its own powers (condition (*)), then $g = (p - 1)p(1 + np)/t$, where n and t are integers, and t divides $p - 1$. The group $\mathfrak{G}$ contains exactly $1 + np$ subgroups of order p, and t is the number of classes of conjugate elements of order p in $\mathfrak{G}$.*

Since g contains the prime p only to the first power, the results of an earlier paper[5] can be applied. For the sake of convenience we mention those facts which will be needed.

The ordinary irreducible representations of $\mathfrak{G}$ are of four different types:
(I) Representations $\mathfrak{A}_\rho$ of a degree $a_\rho = u_\rho p + 1 \equiv 1 \pmod{p}$. Denote by $A_\rho(G)$ the value of the character A_ρ of $\mathfrak{A}_\rho$ for an element G of $\mathfrak{G}$. Then

$$A_\rho(P^i) = 1 \qquad (\text{for } i \not\equiv 0 \pmod{p}). \tag{8, I}$$

II. Representations $\mathfrak{B}_\sigma$ of a degree $b_\sigma = v_\sigma p - 1 \equiv -1 \pmod{p}$. If $B_\sigma(G)$ is the character of $\mathfrak{B}_\sigma$, we have

[5] R. Brauer, *On groups whose order contains a prime number to the first power*, American Journal of Mathematics vol. 54 (1942) part I p. 401, part II, p. 421. I refer to these papers as [1] and [2].

(8, II) $$B_\sigma(P^i) = -1 \qquad \text{(for } i \not\equiv 0 \pmod p\text{)}.$$

(III) Representations $\mathfrak{C}$ of a degree c which is not congruent to $0, 1, -1 \pmod p$ for $t \neq 1$.[6] There exist exactly t such representations $\mathfrak{C}, \mathfrak{C}', \cdots, \mathfrak{C}^{(t-1)}$, and they are algebraically conjugate. The degree c is of the form

$$c = (wp + \delta)/t, \qquad \delta = \pm 1$$

where w is a positive integer. If ϵ is a primitive pth root of unity, suitably chosen, we have for the character $C(G)$ of $\mathfrak{C}$:

(8, III) $$C(P^i) = (-\delta) \sum_{\mu=0}^{q-1} \epsilon^{i\gamma^{\mu t}} \qquad \text{for } i \not\equiv 0 \pmod p.$$

We denote the expression on the right side by $(-\delta)\eta_i$ so that η_i is a Gaussian period of length $q = (p-1)/t$.

(IV) Representation $\mathfrak{D}_\tau$ of a degree $d_\tau = px_\tau \equiv 0 \pmod p$. If $D_\tau(G)$ is the character of $\mathfrak{D}_\tau$, then

(8, IV) $$D_\tau(P^i) = 0 \qquad \text{for } i \not\equiv 0 \pmod p.$$

If we have α representation $\mathfrak{A}_\rho$, $\rho = 1, 2, \cdots, \alpha$, and β representations $\mathfrak{B}_\sigma$, $\sigma = 1, 2, \cdots, \beta$, we have

(9) $$\alpha + \beta = q = (p-1)/t.$$

Furthermore, for elements G of an order prime to p, we have

(10) $$\sum_{\rho=1}^{\alpha} A_\rho(G) + \delta C^{(\nu)}(G) = \sum_{\sigma=1}^{\beta} B_\sigma(G).$$

In particular, for $G = 1$, this gives

(11) $$\sum_\rho a_\rho + \delta c = \sum_\sigma b_\sigma .$$

It is well known that the degrees a_ρ, b_σ, c, d_τ divide the order g of $\mathfrak{G}$ and that g is equal to the sum of the squares of all the degrees, i.e.

(12) $$\sum_\rho a_\rho^2 + \sum_\sigma b_\sigma^2 + tc^2 + \sum_\tau d_\tau^2 = g.$$

It is often convenient to set (as above)

(13) $$a_\rho = u_\rho p + 1, \quad b_\sigma = v_\sigma p - 1, \quad c = (wp + \delta)/t, \quad d_\tau = x_\tau p, \quad (\delta = \pm 1).$$

On substituting these values in (11) and taking (9) into account, we easily obtain

(14) $$\sum_\rho u_\rho + \frac{\delta w + 1}{t} = \sum_\sigma v_\sigma .$$

[6] In the case $t = 1$, $\mathfrak{C}$ can be chosen arbitrarily among the p irreducible representations of degrees not divisible by p. We then choose $\mathfrak{C}$ so that its degree c is of the form $c \equiv -1 \pmod p$. This is always possible. The results given in (III) remain valid for this $\mathfrak{C}$. We then have $\delta = -1$, and $C(P^i)$ is rational.

Substitute the values (13) in (12) and use (9) and (14). A simple computation gives

$$(15) \qquad \sum u_\rho^2 + \sum v_\sigma^2 + \frac{w^2}{t} + \sum x_\tau^2 = \frac{pn - n + 1}{t}.$$

2. Normal subgroups of $\mathfrak{G}$

The number of representations of degree 1 of any group $\mathfrak{G}$ is equal to the index $(\mathfrak{G}:\mathfrak{G}')$ of the commutator-subgroup $\mathfrak{G}'$ of $\mathfrak{G}$. In our case, for $t \neq 1$, $t \neq p - 1$, only the representations $\mathfrak{A}_\rho$ can have degree 1. By (9), their number is at most $(p - 1)/t$. For $t = 1$, we may choose $\mathfrak{C}$ so that its degree is different from 1, and the same argument holds. If $t = p - 1$, then $\alpha + \beta = 1$, cf. (9). Since the 1-representation $G \to 1$ appears among the $\mathfrak{A}_\rho$, we have $\alpha = 1, \beta = 0$, $a_1 = 1$, and (11) gives $c = 1$. Consequently, $\mathfrak{G}$ has exactly p representations of degree p. We thus proved

THEOREM 2. *If the group $\mathfrak{G}$ satisfies the condition (*), then the index $(\mathfrak{G}:\mathfrak{G}')$ of the commutator subgroup $\mathfrak{G}'$ in $\mathfrak{G}$ satisfies the relation*

$$(\mathfrak{G}:\mathfrak{G}') \leqq (p - 1)/t \qquad \textit{if } t \neq p - 1,$$

$$(\mathfrak{G}:\mathfrak{G}') = p, \qquad \textit{if } t = p - 1.$$

If $\mathfrak{G}$ has a normal subgroup $\mathfrak{S}$, then any representation of the factor group $\mathfrak{G}/\mathfrak{S}$ may be considered as a representation of $\mathfrak{G}$. On account of this remark, we prove easily

THEOREM 3. *Let $\mathfrak{G}$ be a group of order g which satisfies condition (*). If $\mathfrak{S}$ is a normal subgroup of an order s divisible by p, then $\mathfrak{S}$ contains the commutator-subgroup $\mathfrak{G}'$ of $\mathfrak{G}$.*

PROOF: Let $\mathfrak{Z}$ be an irreducible representation of $\mathfrak{G}/\mathfrak{S}$ of degree z; let $\zeta(G)$ be the character of the corresponding representation of $\mathfrak{G}$. Since the element P of order p must belong to $\mathfrak{S}$, we have $\zeta(P) = z$. The formulas (8) then show that $z \geqq 2$ is impossible; every irreducible representation of $\mathfrak{G}/\mathfrak{S}$ is of degree 1. Hence $\mathfrak{G}/\mathfrak{S}$ is abelian, i.e. $\mathfrak{S}$ contains $\mathfrak{G}'$, q.e.d.

We now treat normal subgroups of an order which is relatively prime to p. We have

THEOREM 4. *Let $\mathfrak{G}$ be a group of order g which satisfies condition (*). If $\mathfrak{S}$ is a normal subgroup of an order s which is not divisible by p, then s divides $1 + np$ and we have $s \equiv 1 \pmod p$.*[7] *The group $\mathfrak{G}/\mathfrak{S}$ itself satisfies condition (*), and the number t^* of classes containing conjugate elements of order p is the same as the analogous number for $\mathfrak{G}$; i.e. $t = t^*$. The group $\mathfrak{S}$ is contained in the kernel of the representations $\mathfrak{A}_\rho$, $\mathfrak{B}_\sigma$, $\mathfrak{C}^{(\nu)}$ (§1).*

PROOF: The order of $\mathfrak{G}/\mathfrak{S}$ is divisible by p; we have $t^* > 0$. Obviously, $t^* \leqq t$. Consider now the representations of the first p-block of $\mathfrak{G}/\mathfrak{S}$.[8] If

[7] The numbers n and t are defined in theorem 1.

[8] cf. [1], section 8.

$t^* \neq 1$, we find t^* representations whose characters take on distinct algebraically conjugate values for an element of order p. These representations yield t^* representations of $\mathfrak{G}$ with the same property, and the formulas (8) show that $t = t^*$. If $t^* = 1$, we find p representations of $\mathfrak{G}/\mathfrak{S}$ whose characters have non-vanishing rational values for an element of order p. Since this again gives p representations of $\mathfrak{G}$ with the corresponding property, we must have $t = 1$. This shows that $t = t^*$ in any case. The first p-block of $\mathfrak{G}/\mathfrak{S}$ now accounts for $t + (p - 1)/t$ representations of $\mathfrak{G}$ of a degree prime to p. But this is the full number of such representations (cf. §1), and hence $\mathfrak{G}/\mathfrak{S}$ does not contain any other p-block of lowest kind. Then the order of the centralizer of a p-Sylow group of $\mathfrak{G}/\mathfrak{S}$ is equal to p;[9] i.e. $\mathfrak{G}/\mathfrak{S}$ satisfies the condition (*). At the same time we proved that $\mathfrak{S}$ is contained in the kernel of all representations A_ρ, B_σ, $C^{(\nu)}$.

The order g/s of $\mathfrak{G}/\mathfrak{S}$ can be written in the form

$$g/s = (p - 1)p(1 + mp)/t \tag{16}$$

where m is a non-negative integer. Comparison of (16) with (7) shows that $s = (1 + np)/(1 + mp)$. Hence s divides $1 + np$, and we have $s \equiv 1 \pmod p$. This proves theorem 4.

Corollary 1. *Any normal subgroup $\mathfrak{S}$ of $\mathfrak{G}$ of an order prime to p possesses an outer automorphism of order p which leaves only the unit element invariant.*

Proof: Transformation of $\mathfrak{S}$ with an element P of order p in $\mathfrak{G}$ defines such an automorphism.—This shows again that $s \equiv 1 \pmod p$.

Corollary 2. *The kernel of any representation $\mathfrak{B}_\sigma$, $\mathfrak{C}^{(\nu)}$ is the (unique) maximal normal subgroup $\mathfrak{S}^*$ of an order prime to p. The same holds for the kernel of $\mathfrak{A}_\rho$, if the degree a_ρ is not 1.*

Proof: As shown above, $\mathfrak{S}^*$ will belong to each such kernel. But the formulas (8, I), (8, II), (8, III) show that the kernel cannot contain elements of order p, i.e. the kernel itself has an order prime to p, and it coincides therefore with $\mathfrak{S}^*$.

Corollary 3. *We have $\mathfrak{S}^* \subseteq \mathfrak{G}'$. If $t \neq p - 1$, the group $\mathfrak{G}'/\mathfrak{S}^*$ is simple. If $t = p - 1$, $\mathfrak{G}' = \mathfrak{S}^*$.*

Proof: The group $\mathfrak{G}'$ can be defined as the intersection of the kernels of the representations of degree 1. Theorem 4 then gives $\mathfrak{S}^* \subseteq \mathfrak{G}'$. If $t \neq p - 1$, the order of $\mathfrak{G}'$ is divisible by p (cf. theorem 2). From theorem 3 and the definition of $\mathfrak{S}^*$ it follows that no normal subgroup of $\mathfrak{G}$ lies between $\mathfrak{G}'$ and $\mathfrak{S}^*$. Then the group $\mathfrak{G}'/\mathfrak{S}^*$ is a minimal normal subgroup of $\mathfrak{G}/\mathfrak{S}^*$, and hence $\mathfrak{G}'/\mathfrak{S}^*$ is a direct product of isomorphic simple groups. But since the order of $\mathfrak{G}'/\mathfrak{S}^*$ contains p to the first power, this implies that $\mathfrak{G}'/\mathfrak{S}^*$ is simple. If $t = p - 1$, theorem 2 shows that $\mathfrak{G}' = \mathfrak{S}^*$.

Corollary 4. *If n is not of the form $n = u + m + ump$ (u, m positive integers), in particular if $n < p + 2$, then $\mathfrak{G}$ does not contain a normal subgroup $\mathfrak{S} \neq \{1\}$ of an order s smaller than $1 + np$.*

[9] cf. [1], theorem 1.

PROOF: If $s \equiv 0 \pmod p$, theorems 2 and 3 give $s \geqq (g:(p-1)/t) = p(1+np)$. If $s \not\equiv 0$, theorem 4 shows that s is of the form $1 + up$ where u is a positive integer. From (7) and (16) we obtain

$$1 + np = (1 + mp)(1 + up).$$

Hence $n = u + m + ump$. Under our present assumption, we must have $m = 0$, i.e. $s = 1 + np$ and this proves the corollary.

We now distinguish two cases:

CASE I. *The Group* $\mathfrak{G}$ *contains a normal subgroup of order* $1 + np$.

CASE II. *The group* $\mathfrak{G}$ *does not contain a normal subgroup of order* $1 + np$. In other words, in case I the order s^* of $\mathfrak{S}^*$ is equal to $1 + np$ while in case II s^* is smaller than $1 + np$.

THEOREM 5. *We have case I, if and only if one of the following two sets of conditions holds*

(a) $$t = p - 1.$$

(b) $t < p - 1$, *and the first and second commutator subgroups* $\mathfrak{G}'$ *and* $\mathfrak{G}''$ *of* $\mathfrak{G}$ *are different.*

In case (a), $\mathfrak{G}'$ *has the order* $1 + np$, *and* $\mathfrak{G}/\mathfrak{G}'$ *is cyclic of order* p. *In case (b), the group* $\mathfrak{G}''$ *has the order* $1 + np$ *and* $\mathfrak{G}'$ *has the order* $p(1 + np)$; $\mathfrak{G}/\mathfrak{G}''$ *is metacyclic and can be defined by the equations* (5).

PROOF: The case $t = p - 1$ is trivial, cf. theorem 2 and (7); we may assume $t < p - 1$. If $\mathfrak{S}^*$ is an invariant subgroup of order $1 + np$ in $\mathfrak{G}$, then $\mathfrak{G}/\mathfrak{S}^*$ is a group of order $p(p-1)/t$, which satisfies condition (*) and in which t classes of conjugate elements contain elements of order p. Hence $\mathfrak{G}/\mathfrak{S}^*$ contains a subgroup of type (5), and since this subgroup has order $p(p-1)/t$, the group $\mathfrak{G}/\mathfrak{S}^*$ itself is a metacyclic group of type (5). In particular, $\mathfrak{G}/\mathfrak{S}^*$ contains a normal subgroup $\mathfrak{G}_1/\mathfrak{S}^*$ of index $(p-1)/t$. Then $\mathfrak{G}_1$ is a normal subgroup of index $(p-1)/t$ of $\mathfrak{G}$, and theorems 2 and 3 now show that $\mathfrak{G}_1 = \mathfrak{G}'$. We may apply theorem 2 to $\mathfrak{G}'$ which again satisfies condition (*). Since $\mathfrak{G}'$ contains a normal subgroup $\mathfrak{S}^*$ of index p, this group $\mathfrak{S}^*$ must be the commutator subgroup $\mathfrak{G}''$ of $\mathfrak{G}'$. Conversely, assume that $\mathfrak{G}' \neq \mathfrak{G}''$. According to theorem 2 we have $(\mathfrak{G}:\mathfrak{G}') \leqq (p-1)/t$. The order of $\mathfrak{G}'$ then is divisible by p, and $\mathfrak{G}'$ also satisfies the condition (*). If the index $(\mathfrak{G}':\mathfrak{G}'')$ was prime to p, the group $\mathfrak{G}''$ would have an order divisible by p, and theorem 3 would give $\mathfrak{G}'' \supseteq \mathfrak{G}'$, i.e. $\mathfrak{G}'' = \mathfrak{G}'$. Hence $(\mathfrak{G}':\mathfrak{G}'')$ is divisible by p. Now theorem 2, applied to $\mathfrak{G}'$, gives $(\mathfrak{G}':\mathfrak{G}'') = p$ and, therefore, $(\mathfrak{G}:\mathfrak{G}'') \leqq p(p-1)/t$. However, theorem 4 shows that the order of the normal subgroup $\mathfrak{G}''$ of $\mathfrak{G}$ must divide $1 + np$. As $\mathfrak{G}$ has the order $p(p-1)(1+np)/t$, we now see that $\mathfrak{G}''$ has the order $1 + np$. This completes the proof of theorem 5.

COROLLARY 5. *In case II, the order* g' *of the group* $\mathfrak{G}'$ *is given by*

(17) $$g' = (p-1)p(1+np)/t'$$

where n is the same number as in (7) and t' denotes the number of classes of conjugate elements in $\mathfrak{G}'$ which contain elements of order p. We have $t \mid t'$, $t' \mid (p - 1)$, and $t \leqq t' < p - 1$. Furthermore, $\mathfrak{G}' = \mathfrak{G}''$. The group $\mathfrak{G}/\mathfrak{G}'$ is cyclic.

PROOF: It follows from theorem 5 that $t < p - 1$ and that $\mathfrak{G}' = \mathfrak{G}''$. The group $\mathfrak{G}'$ also satisfies condition (*), and since it contains all subgroups of order p of $\mathfrak{G}$, we obtain (17). The number t divides t', because g' divides g. If we had $t' = p - 1$, theorem 2 would give $\mathfrak{G}' \neq \mathfrak{G}''$. The element Q in (5) has the property that its (t'/t)th power is the first power which belongs to $\mathfrak{G}'$. This shows that $\mathfrak{G}/\mathfrak{G}'$ is cyclic.

COROLLARY 6. *If n is not of the form $n = u + m + ump$, (u, m positive integers), in particular if $n < p + 2$, then $\mathfrak{G}'$ is simple in case II.*

PROOF: Corollary 4 shows that $\mathfrak{S}^* = \{1\}$, and corollary 3 now proves the statement.

Finally, we treat the three kinds of groups mentioned in the introduction. We prove

THEOREM 6. *If $\mathfrak{G}$ is a transitive permutation group of degree p (p a prime number), then $\mathfrak{G}$ does not contain any normal subgroup of an order prime to p and different from $\{1\}$; the group $\mathfrak{G}'$ is simple (or of order 1). If $\mathfrak{G}$ is a doubly transitive group of degree $p + 1$ or if $\mathfrak{G}$ is an irreducible linear group with center $\{1\}$ in a p-dimensional space, then any normal subgroup of an order prime to p is abelian. In all these cases, the composition series of $\mathfrak{G}$ has at most one non-cyclic factor group.*

PROOF: If $\mathfrak{G}$ is a transitive permutation group of degree p, then $\mathfrak{G}$ possesses a reducible (1–1)-representation $\mathfrak{Z}$ of degree p, whose character has the value 0 for elements of order p. As shown by the formulas (8, I), this representation cannot consist of representations $\mathfrak{A}_\rho$ exclusively; furthermore it cannot contain any constitutent $\mathfrak{D}_\tau$. Theorem 4 and corollary 2 now show that $\mathfrak{Z}$ has the kernel $\mathfrak{S}^*$. However, $\mathfrak{Z}$ was a (1–1)-representation. We thus find $\mathfrak{S}^* = \{1\}$, and corollary 3 shows that $\mathfrak{G}'$ is simple (or $\mathfrak{G}' = \{1\}$, if $g = p$).

Any doubly transitive permutation group $\mathfrak{G}$ of degree $p + 1$ possesses an irreducible (1–1)-representation of degree p. As condition (*) holds for any such $\mathfrak{G}$, we easily see that $\mathfrak{G}$ has the center $\{1\}$. It is, therefore, sufficient to treat irreducible linear groups with the center $\{1\}$ in a p-dimensional space. It follows here that $\mathfrak{S}^*$, considered as a linear group, must be reducible since the dimension does not divide the order s^*. Since $\mathfrak{S}^*$ is a normal subgroup, it splits into constituents of the same degree z. Then $z = 1$, and $\mathfrak{S}^*$ is abelian.

The last statement of theorem 6 follows from corollary 3.

COROLLARY 7. *If $\mathfrak{G}$ is a primitive irreducible linear group in a p-dimensional space and if $\mathfrak{G}$ has the center $\{1\}$, then $\mathfrak{G}'$ is simple.*

PROOF: When $\mathfrak{G}$ is primitive, the normal abelian subgroup $\mathfrak{S}^*$ must lie in the center. We then have $\mathfrak{S}^* = \{1\}$ under our assumptions. Now corollary 3 gives the statement.

There is a well known theorem of Burnside which states that a transitive permutation group $\mathfrak{G}$ of degree p is either doubly transitive or it is metacyclic of the type (5). With the methods used here, this could be proved in the fol-

lowing manner. If $\mathfrak{G}$ is not doubly transitive, the permutation representation $\mathfrak{Z}$ splits into the 1-representation $\mathfrak{A}_1$ and at least two more constitutents not all of which can be of type $\mathfrak{A}_\rho$. Then one constituent at least is a $\mathfrak{C}^{(\nu)}$. Since the character is rational, all the conjugate $\mathfrak{C}^{(\nu)}$ appear, and we find

$$\mathfrak{Z} = \mathfrak{A}_1 + \sum_{\nu=0}^{t-1} \mathfrak{C}^{(\nu)}, \qquad c = (p-1)/t, \qquad t \geqq 2.$$

If $t > 2$, the group $\mathfrak{G}$ must have a normal subgroup of order p,[10] and therefore $\mathfrak{G}$ is of type (5). If $t = 2$, we have also to consider the case that $\mathfrak{G} \cong LF(2, p)$. It is not very difficult to exclude this last possibility. However, the proof may be omitted.

3. Conditions for the degrees a_ρ, b_σ, c

We now make use of the fact that the degrees a_ρ, b_σ, c (cf. (13)), divide the order $g = (p-1)p(1+np)/t$ of $\mathfrak{G}$. For the a_ρ, we certainly must have $(u_\rho p + 1) \mid (p-1)(1+np)$. If the sign δ has the value $+1$, the condition for c gives $(wp+1) \mid (p-1)(1+np)$. The question we have to treat then is this: When is

$$(up+1) \mid (p-1)(1+np)$$

(where $u = u_\rho$ or $u = w$)? Set $up + 1 = m_1m_2$ with $m_1 \mid (p-1)$ and $m_2 \mid (1+np)$. Then $p \equiv 1$, $up + 1 \equiv 0 \pmod{m_1}$ and hence $u + 1 \equiv 0 \pmod{m_1}$. Further, $up + 1 \equiv 0$, $1 + np \equiv 0 \pmod{m_2}$, and this gives $(n-u)p \equiv 0$ and hence $n - u \equiv 0 \pmod{m_2}$. It now follows that $up + 1 = m_1m_2$ divides $(u+1)(n-u)$. We may set

$$(u+1)(n-u) = h(up+1) \tag{18}$$

with an integral h. Then $h(up+1) \equiv 0 \pmod{u+1}$. Since $u \equiv -1 \pmod{u+1}$, this gives

$$(u+1) \mid h(p-1). \tag{19}$$

Assume first that $h = -h' < 0$. Then $(u+1)(u-n) = h'(up+1)$ which gives

$$u^2 = u(h'p + n - 1) + h' + n > u(h'p + n - 1).$$

Hence $u > h'p + n - 1$ while (19) yields $h'(p-1) \geqq u + 1$, i.e. $u \leqq h'p - h' - 1 \leqq h'p + n - 1$. This is a contradiction, the case $h < 0$ is impossible. From (18) it follows that $n = (hup + u^2 + u + h)/(u+1)$. We therefore have

LEMMA 1. *Denote by $F(p, u, h)$ the rational function*

$$F(p, u, h) = \frac{hup + u^2 + u + h}{u+1} \tag{20}$$

---

[10] cf. [2], theorem 3.

of p, u and h. If $gt = (p-1)p(1+np)$ is divisible by a number $1+up$ where u is a non-negative integer, then there exists an integer $h \geqq 0$ such that

$$n = F(p, u, h) \tag{21}$$

and that $(u+1) \mid h(p-1)$.

In a similar fashion, we have to find the condition that $vp - 1$ divides $gt = (p-1)p(1+np)$ where $v > 0$, ($v = v_\sigma$, or $v = w$ for $\delta = -1$). We set $vp - 1 = m_1 m_2$ with $m_1 \mid (p-1)$ and $m_2 \mid (1+np)$ and derive $v - 1 \equiv 0 \pmod{m_1}$, $v + n \equiv 0 \pmod{m_2}$. Hence

$$(v-1)(v+n) = h(vp-1) \tag{22}$$

for some integer $h \geqq 0$. For fixed n, p, h, this is a quadratic equation for v. The second root $v' = (h-n)/v$ must be integral too. For $h = 0$, we have $v' = -n/v \leqq 0$. If $h \neq 0$, replace v by 1 in (22). The left side of (22) then vanishes while the right side is positive. This again gives $v' \leqq 0$.

Set $u = -v' = (n-h)/v$; then $u \geqq 0$. Since v' satisfies the equation (22), this equation remains correct when v is replaced by $-u$. We thus come back to (18). Since (18) implied (19) and (21), we have proved

LEMMA 2. *If $gt = (p-1)p(1+np)$ is divisible by $vp - 1$, where v is a positive integer, then there exists a non-negative integer h such that $u = (n-h)/v$ is integral and not negative, and that the relations hold*

$$n = F(p, u, h); \qquad (u+1) \mid h(p-1). \tag{23}$$

If $u = 0$, then $n = h$, and (22) gives $v = pn - n + 1$. However, it follows from (15) that $v_\sigma = pn - n + 1$ or $w = pn - n + 1$ are possible only when $v_\sigma = 1$ or $w = 1$, i.e. when $n = 0$. If $h = 0$, then $u = n$ and $v = 1$.

THEOREM 7. *If $\mathfrak{G}$ is a group satisfying the condition (*), then find all representations of n in the form $n = F(p, u^{(\nu)}, h^{(\nu)}) = (h^{(\nu)}u^{(\nu)}p + u^{(\nu)2} + u^{(\nu)} + h^{(\nu)})/(u^{(\nu)} + 1)$ with positive integers $u^{(\nu)}$, $h^{(\nu)}$. The degrees of the irreducible representations of $\mathfrak{G}$, as far as they are prime to p can only have some of the values*

$$a_\rho = 1, \qquad a_\rho = u^{(\nu)}p + 1, \qquad a_\rho = np + 1. \tag{24}$$

$$b_\sigma = p - 1, \qquad b_\sigma = v^{(\nu)}p - 1. \tag{25}$$

$$c = (np+1)/t, \qquad c = (u^{(\nu)}p + 1)/t, \qquad c = (p-1)/t, \tag{26}$$

$$c = (v^{(\nu)}p - 1)/t$$

where $v^{(\nu)}$ is set equal to $(n - h^{(\nu)})/u^{(\nu)}$.

For $h > 0$ and variable u, we have $(\partial/\partial u)F(p, u, h) > 0$. Since we are only interested in solutions u, h of $n = F(p, u, h)$ with $1 \leqq u \leqq h(p-1) - 1$, we must have

$$F(p, 1, h) = \frac{hp + h + 2}{2} \leqq n \leqq F(p, h(p-1) - 1, h) = 2ph - h - 2. \tag{27}$$

This gives

THEOREM 8. *In theorem 7, only values $h = h^{(\nu)}$ have to be considered for which*

$$\frac{n+2}{2p-1} \leqq h \leqq \frac{2n-2}{p+1}. \tag{28}$$

To each $h^{(\nu)}$ there belongs at most one $u^{(\nu)}$, and this $u^{(\nu)}$ satisfies the conditions $u^{(\nu)} \mid (n - h^{(\nu)})$, $(u^{(\nu)} + 1) \mid h^{(\nu)}(p - 1)$.

The last remark follows from the fact that the equation $n = F(p, u, h)$ is equivalent to (18), and since $h < n$, we have one positive root u. Unless n has one of the following values

$$n = \begin{array}{lllll} \frac{p+3}{2}, & \frac{2p+7}{3}, & \frac{3p+13}{4}, & \cdots & (h = 1; u = 1, 2, 3, \cdots) \\ p+2, & \frac{4p+8}{3}, & \frac{3p+7}{2}, & \cdots & (h = 2; u = 1, 2, 3, \cdots) \\ \frac{3p+5}{2}, & 2p+3, & \frac{9p+15}{4}, & \cdots & (h = 3; u = 1, 2, 3, \cdots) \\ \cdots & & & & \cdots \end{array} \tag{29}$$

only the values

$$a_\rho = 1, \quad a_\rho = np + 1, \quad b_\sigma = p - 1, \quad c = (np + 1)/t, \quad c = (p - 1)/t \tag{30}$$

are possible in theorem 7.

As an example for theorems 7 and 8, we choose the case $t = 2$, $p = 11$, $n = 157$ for which $g = 95040$.[11] We find here $9 \leqq h \leqq 26$. Actually, only the values $h = 10, 12, 14, 17, 26$ give integral values for u. From (15), we obtain $u_\rho \leqq 28$, $v_\sigma \leqq 28$, $x_\tau \leqq 28$, $w \leqq 39$. We thus obtain as the possible values of the degrees a_ρ and b_σ:

$$a_\rho = 1, 12, 45, \text{or } 144; \qquad b_\sigma = 10, 32, 54, \text{or } 120,$$

while c has one of the values 5, 6, 16, 27, 60, 72, 160, 190. The sign $\delta = \pm 1$ is determined by $2c \equiv -\delta \pmod{11}$. The total number of degrees a_ρ and b_σ is five, and their values together with c must satisfy the equation (11). If we further assume that $\mathfrak{G}$ is simple, only one of the a_ρ has the value 1. On the basis of these remarks and using further properties of group representations, it is possible to compute the full table of group characters of $\mathfrak{G}$.

---

[11] This case is relatively complicated as n will appear five times in the table (29) for $p = 11$. The case has been chosen because there exists a simple group $\mathfrak{G}$ of order 95040, the five times transitive permutation group of degree 12 of Mathieu. It is possible to show that, for any simple group of order 95040 and $p = 11$, the condition (*) must hold and that t must have the value 2. The characters of the Mathieu groups have been obtained by Frobenius (Sitzungsberichte der Preussischen Akademie der Wissenschaften, Berlin 1903). Our result shows that any simple group of order 95040 has the same table of characters as the Mathieu group. This seems to make it appear highly probable that only one simple group of order 95040 exists.

4. Relations between the characters of $\mathfrak{G}$ and those of $\mathfrak{N} = \mathfrak{N}(\mathfrak{P})$

Every representation of $\mathfrak{G}$ defines a new representation of any given subgroup. We apply this for the subgroup $\mathfrak{N} = \mathfrak{N}(\mathfrak{P})$ defined by (5). In this manner, we can obtain some new information about the characters of $\mathfrak{G}$ from the formulas of section 1, and this will be needed later.

It is easy to find all irreducible characters of the group $\mathfrak{N}$ of order $pq = p(p-1)/t$.[12] Let ω be a primitive qth root of unity. We then have q linear characters ω_μ, $(\mu = 0, 1, 2, \cdots, q-1)$ defined by

$$(31) \qquad \omega_\mu(Q^j) = \omega^{\mu j}, \qquad \omega_\mu(P^j) = 1.$$

Besides, we have t conjugate characters $Y^{(\nu)}$ of degree q. We only notice that

$$(32) \qquad Y^{(\nu)}(Q^j) = 0 \qquad \text{for } j \not\equiv 0 \pmod{q}.$$

Set

$$\Omega = \omega_0 + \omega_1 + \cdots + \omega_{q-1}.$$

Then, we have

$$(33) \qquad \Omega(1) = \Omega(P^i) = q, \qquad \Omega(Q^j) = 0 \qquad \text{for } j \not\equiv 0 \pmod{q}.$$

Every element N of $\mathfrak{N}$ which is not a power of P is conjugate to some Q^j and hence $\Omega(N)$ vanishes.

The expressions $A_\rho(N)$, $B_\sigma(N)$, $C^{(\nu)}(N)$, $D_\tau(N)$, (N in $\mathfrak{N}$), form (reducible) characters of $\mathfrak{N}$. Hence they can be expressed by the $\omega_\mu(N)$ and the $Y^{(\nu)}(N)$. From (33), (8) and (13), we obtain (N in the sums ranging over the elements of $\mathfrak{N}$)

$$\begin{aligned}
\sum A_\rho(N)\Omega(N) &= \sum_{i=0}^{p-1} A_\rho(P^i)\Omega(P^i) = q(a_\rho + p - 1) = qp(u_\rho + 1);\\
\sum B_\sigma(N)\Omega(N) &= \sum B_\sigma(P^i)\Omega(P^i) = q(b_\sigma - p + 1) = qp(v_\sigma - 1);\\
\sum C^{(\nu)}(N)\Omega(N) &= \sum C^{(\nu)}(P^i)\Omega(P^i) = q\left(c - \delta\sum_{i=1}^{p-1}\eta_i\right) = q(c + \delta q)\\
&\qquad = q(wp + \delta + \delta tq)/t = qp(w + \delta)/t;\\
\sum D_\tau(N)\Omega(N) &= \sum D_\tau(P^i)\Omega(P^i) = qd_\tau \qquad = pqx_\tau.
\end{aligned}$$

From the orthogonality relations for the characters of $\mathfrak{N}$, we now obtain

LEMMA 3. *If we consider the characters of $\mathfrak{G}$ only for elements N of the subgroup $\mathfrak{N}$, then $A_\rho(N)$ contains $u_\rho + 1$ of the $\omega_\mu(N)$, $B_\sigma(N)$ contains $v_\sigma - 1$ of the $\omega_\mu(N)$, $C^{(\nu)}(N)$ contains $(w + \delta)/t$ of the ω_μ, and $D_\tau(N)$ contains x_τ of the $\omega_\mu(N)$. The same $\omega_\mu(N)$ appear in the different $C^{(\nu)}(N)$, $\nu = 0, 1, 2, \cdots, t-1$.*

The last remark follows easily from the fact that $C(G)$ can be carried into

[12] The equations (31) and (32) can easily be derived from our general results applied to the case $n = 0$.

each $C^{(\nu)}(G)$ by a change in the choice of the primitive pth root of unity ϵ, because this change does not affect ω_μ. We also show

LEMMA 4. *We have*

$$\left\{\begin{aligned}\sum_\rho a_\rho A_\rho(N) + \sum_\sigma b_\sigma B_\sigma(N) + c\sum_\nu C^{(\nu)}(N) + \sum_\tau d_\tau D_\tau(N)\\ = (1+pn)\sum_{\mu=0}^{q-1}\omega_\mu(N) + \cdots .\end{aligned}\right. \tag{34}$$

$$\sum_\rho A_\rho(N) + \delta C(N) = \sum_\sigma B_\sigma(N) + \sum_{\mu=0}^{q-1}\omega_\mu(N) + \cdots . \tag{35}$$

$$\begin{aligned}\sum_\rho u_\rho A_\rho(N) + \sum_\sigma v_\sigma B_\sigma(N) + wC(N) + \sum_\tau x_\tau D_\tau(N)\\ = n\sum_{\mu=0}^{q-1}\omega_\mu(N) + \cdots .\end{aligned} \tag{36}$$

where the dots stand for a linear homogeneous combination of the $Y^{(\nu)}(N)$.

PROOF: The left side in (34) is the character $R(N)$ of the regular representation of $\mathfrak{G}$, and we have $R(1) = g$, $R(N) = 0$ for $N \neq 1$. Then (34) follows from the orthogonality relations and (31).

The expression

$$S(N) = \sum_\rho A_\rho(N) + \delta C(N) - \sum_\sigma B_\sigma(N)$$

can be written as a linear combination of the characters of $\mathfrak{N}$ with integral rational coefficients. From (8), (9) and (10), we obtain

$$\begin{aligned}\sum_{i=1}^{p-1} S(P^i) &= \sum_\rho (p-1) + \delta\left(-\delta\sum_{i=1}^{p-1}\eta_i\right) - \sum_\sigma(-p+1)\\ &= \alpha(p-1) + q + \beta(p-1) = pq;\end{aligned}$$

$$S(N) = 0 \quad \text{for} \quad N \neq P, P^2, \cdots, P^{p-1}; \qquad N \text{ in } \mathfrak{N}.$$

Then (31) gives

$$\sum S(N)\bar{\omega}_\mu(N) = pq$$

which shows that $S(N)$ contains ω_μ with the coefficient 1. Finally, (36) is obtained by subtracting (35) from (34) and dividing by p, taking into account that $C^{(\nu)}(N)$ and $C(N)$ contain the same ω_μ.

As a first application of the considerations of this section we prove

THEOREM 9.[13] *Let $\mathfrak{G}$ be a group which satisfies condition (*). If $\mathfrak{G}$ possesses an irreducible representation $\mathfrak{Z}$ of degree $p-1$, then either the number t is even or the index of the commutator-group $\mathfrak{G}'$ in $\mathfrak{G}$ is even.*

PROOF: It follows from lemma 3 that the character $\zeta(N)$ of $\mathfrak{Z}(N)$ (N in $\mathfrak{N}$),

---

[13] If $p = 2$, then $t = 1 = p - 1$, and the theorem follows from Theorem 2.

contains only characters $Y^{(\nu)}(N)$ and no $\omega_\mu(N)$. We then must have $(p-1)/q = t$ such constituents $Y^{(\nu)}(N)$. It follows from (32) that $\mathfrak{Z}(Q)$ has the characteristic roots $1, \omega, \cdots, \omega^{q-1}$, each taken t times. Hence the determinant of $\mathfrak{Z}(Q)$ has the value $(-1)^{(q+1)t} = (-1)^{p-1+t} = (-1)^t$. The determinant of $\mathfrak{Z}(G)$, G in $\mathfrak{G}$, forms a representation of degree 1 of $\mathfrak{G}$. If t is odd, an even power of this representation will give the 1-representation of G. This implies that $(\mathfrak{G}:\mathfrak{G}')$ is even.

5. Proof of the main theorem

In what is to follow we shall assume that $\mathfrak{G}$ satisfies the following further condition

(**) *The commutator-subgroup of $\mathfrak{G}$ is equal to $\mathfrak{G}$.*

In the notation of section 2, we must have case II.[14] This shows that groups $\mathfrak{G}$ are excluded which contain a normal subgroup $\mathfrak{H}$ such that $\mathfrak{G}/\mathfrak{H}$ is metacyclic of order $p(p-1)/t$. Conversely, when $\mathfrak{G}$ is any group which satisfies the condition (*) and falls under case II, its commutator-subgroup $\mathfrak{G}'$ satisfies conditions (*) and (**). The number n is the same for $\mathfrak{G}$ and $\mathfrak{G}'$ and $\mathfrak{G}$.is obtained from $\mathfrak{G}'$ by a cyclic extension. We now prove

THEOREM 10. *Let $\mathfrak{G}$ be a group which contains an element P of prime order p which commutes only with its own powers and assume that $\mathfrak{G}$ is equal to its commutator-subgroup $\mathfrak{G}'$. If $g = (p-1)p(1+np)/t$ is the order of $\mathfrak{G}$ where $1 + np$ is the number of conjugate subgroups of order p in $\mathfrak{G}$, then the number n must be of the form*

$$n = \frac{puh + u^2 + u + h}{u + 1} \tag{37}$$

where.u and h are positive integers, except in the following two cases

(a) $n = 1, t = 2$. *Here,* $\mathfrak{G} \cong LF(2, p)$, $(p > 3)$.

(b) $n = \dfrac{p-3}{2}$, $t = \dfrac{p-1}{2}$, $p = 2^\mu + 1 > 3$ *a Fermat prime. Here,* $\mathfrak{G} \cong LF(2, p-1)$.

COROLLARY. *If $n < (p+3)/2$, then $\mathfrak{G}$ must be either of the type (a) or of the type (b).*

PROOF: Suppose that n is not representable in the form (37). Then the degrees of the irreducible representations of $\mathfrak{G}$ are either divisible by p or have one of the values (30). Because of condition (**) the degree 1 appears only once, say $a_1 = 1$. If t was odd, theorem 9 shows that the degree $p - 1$ does not appear. It follows from (11) that this is impossible. Hence

$$t \equiv 0 \pmod 2. \tag{38}$$

The degree $(p-1)/t$ is impossible for $t > 2$;[15] for $t = 2$ it occurs only in the case $\mathfrak{G} \cong LF(2, p)$, i.e. in the case (a). Hence we may exclude this possibility.

[14] This implies that $p \neq 2$.

[15] If n is not of the form (37), it is not of the form $n = ump + u + m$ with positive integers u and m, since otherwise we could set $h = (u+1)m$ and would obtain a representation (37). Then corollary 6 (section 2) shows that $\mathfrak{G}'$ is simple. Because of conditions (**), $\mathfrak{G}$ is simple. Now, [2], theorems 3 and 4, can be applied.

We now see that we must have

$$(39)\qquad \begin{aligned} &a_1 = 1, \qquad a_2 = a_3 = \cdots = a_\alpha = 1 + np, \\ &b_1 = b_2 = \cdots = b_\beta = p - 1, \qquad c = (pn + 1)/t \end{aligned}$$

and the sign δ has the value $+1$. The values of α and β can be obtained from (9), (13), (14), and (39) which give

$$(40)\qquad \alpha + \beta = (p - 1)/t = q,$$

$$(41)\qquad (\alpha - 1)n + \frac{n + 1}{t} = \beta.$$

In particular, $n + 1$ is divisible by t; we set

$$(42)\qquad n + 1 = st,$$

and then obtain

$$(43)\qquad 1 + np = 1 + (qt + 1)(st - 1) = tr$$

where

$$(44)\qquad r = qst + s - q = ps - q.$$

By (43), the order g of $\mathfrak{G}$ can be written in the form

$$(45)\qquad g = (p - 1)pr = qtpr.$$

The next step is to show that r and $p - 1 = qt$ are relatively prime. Using (44), (42), (41), (40), we obtain successively

$$s \equiv 0, \qquad n \equiv -1, \qquad (\alpha - 1)(-1) + s \equiv \beta, \qquad \alpha + \beta \equiv 0 \qquad (\mathrm{mod}\ (r, q))$$

whence it follows that $1 \equiv 0$, i.e. that $(r, q) = 1$. In a similar manner, we have

$$s \equiv q, \qquad n \equiv -1, \qquad (\alpha - 1)(-1) + s \equiv \beta, \qquad \alpha + \beta \equiv s \quad (\mathrm{mod}\ (r, t))$$

which gives $s + 1 \equiv s$ (mod (r, t)), i.e. $(r, t) = 1$. Hence we have

$$(46)\qquad (r, p - 1) = 1.$$

From (45) and (46), it follows that for any prime l dividing $p - 1$ the characters B_σ are of highest kind.[16] This implies

$$(47)\qquad B_\sigma(L) = 0$$

for elements L of $\mathfrak{G}$ whose order is divisible by l. For the primes m dividing r the characters of degrees $1 + pn = rt$ and $(1 + pn)/t = r$ are of the highest kind. Hence

$$(48)\qquad A_\rho(M) = 0 \quad \text{for} \quad \rho \neq 1, \qquad C^{(\nu)}(M) = 0$$

---

[16] Cf. R. Brauer and C. Nesbitt, Annals of Mathematics, vol. 42, p. 556 (1941), Chapter II.

for elements M of $\mathfrak{G}$ whose order is divisible by m. Because of the assumption (*), the order of the elements L and M is not divisible by p, and hence the equation (10) holds for these elements. If an element G of $\mathfrak{G}$ would be an element L and an element M at the same time, then every term in (10) except $A_1(G) = 1$ would vanish, and this is impossible. Hence the elements of $\mathfrak{G}$ are distributed into four disjoint sets: (I) The 1-element, (II) the elements of order p, (III) the elements L whose order is divisible by at least one prime factor of $p - 1$, (IV) the elements M whose order is divisible by at least one prime factor of r.

Consider now the following element of the group ring Γ belonging to $\mathfrak{G}$

$$T = \sum_{j=1}^{q-1} Q^j. \tag{49}$$

We wish to show that (ρ now will always denote one of the values $2, 3, \cdots, \alpha$)

$$\begin{aligned} &A_1(T) = q - 1, \qquad A_\rho(T) = -(n + 1), \qquad B_\sigma(T) = 0, \\ &\qquad\qquad C^{(\nu)}(T) = -(n + 1)/t, \qquad D_\tau(T) = (q - 1)x_\tau^{*}. \end{aligned} \tag{50}$$

The proof of (50) can be obtained from the results of section 4. By (39) and (13) we have

$$u_1 = 0, \qquad u_2 = \cdots = u_\alpha = n, \qquad v_1 = \cdots = v_\beta = 1, \qquad w = n, \qquad \delta = 1.$$

Lemma 3 shows that $A_1(N)$ contains exactly one ω_μ, $A_\rho(N)$ with $\rho > 1$ contains exactly $n + 1$ of the ω_μ, $B(N)$ does not contain any ω_μ, $C(N)$ contains exactly $(n + 1)/t$ of the ω_μ, $D_\tau(N)$ contains x_τ of the ω_μ. Here, N is an element of $\mathfrak{N} = \mathfrak{N}(\mathfrak{P})$, cf. (5). It now follows from (35) that each of the ω_μ appears in one of the characters $A_1(N), A_2(N), \cdots, A_\alpha(N), C(N)$, while (36) shows that the ω_μ appearing in $A_\rho(N)$, $\rho > 1$, do not occur in any other character, and that the ω_μ appearing in $C(N)$ occur only in the $C^{(\nu)}(N)$. Since, obviously, $A_1(N) = 1 = \omega_0(N)$, this implies that the $A_\rho(N)$ with $\rho > 1$ and $C(N)$ contain only ω_μ for which $\mu \geqq 1$ whereas the D_τ contain only $\omega_0(N)$.

The elements Q^j belong to $\mathfrak{N}$. On account of (31) and (32), we have

$$\omega_0(T) = q - 1, \qquad \omega_\mu(T) = -1 \quad \text{for} \quad \mu \neq 0, \qquad Y^{(\nu)}(T) = 0.$$

The first formula (50) is obvious, as $A_1(N) = 1 = \omega_0(N)$. Since $A_\rho(N)$ for $\rho > 1$ is a sum of $n + 1$ terms ω_μ with $\mu > 0$ and of terms $Y^{(\nu)}$, we find $A_\rho(T) = -(n + 1)$. The remaining formulas (50) follow in the same manner using the facts that, apart from terms $Y^{(\nu)}$, the character $C^{(\nu)}(N)$ is a sum of exactly $(n + 1)/t$ terms ω_μ with $\mu > 0$, the character $D_\tau(N)$ contains only $x_\tau\omega_0$, and the characters $B_\sigma(N)$ do not contain any term except terms $Y^{(\nu)}$.

Let ζ range over all the characters $A_1, A_\rho, B_\sigma, C^{(\nu)}, D_\tau$ of $\mathfrak{G}$. If L again is an element whose order contains at least one prime factor of $p - 1$, then (47) and (50) give

$$\begin{aligned}(51)\qquad \sum \zeta(T)\zeta(L) = (q-1) - (n+1)\sum_{\rho=2}^{\alpha} A_\rho(L) \\ - (n+1)C(L) + (q-1)\sum_{\tau} x_\tau D_\tau(L)\end{aligned}$$

since the t characters $C^{(\nu)}$ have the same value for the element L.

In order to compute these expressions in a different manner, we use the orthogonality relations for group characters. If ζ again ranges over all characters of $\mathfrak{G}$, we have

$$\begin{aligned}(52)\qquad 1 + \sum_{\rho=2}^{\alpha} A_\rho(G)A_\rho(H) + \sum_{\sigma} B_\sigma(G)B_\sigma(H) \\ + \sum_{\nu} C^{(\nu)}(G)C^{(\nu)}(H) + \sum_{\tau} D_\tau(G)D_\tau(H) \\ = \sum \zeta(G)\zeta(H) = n(G)\delta(G, H)\end{aligned}$$

where $n(G)$ is the order of the normalizer of $\mathfrak{G}$, and where $\delta(G, H)$ has the value 1 or 0 according as the elements G and H^{-1} of $\mathfrak{G}$ are or are not conjugate.

In particular, set $G = 1$ and $H = L$. The value $\zeta(1)$ is equal to the degree of the character and can be found from (39) and (13). Again we use (47) and the fact that $C^{(\nu)}(L) = C(L)$. Thus

$$1 + (1+np)\sum_{\rho} A_\rho(L) + (1+np)C(L) + p\sum_{\tau} x_\tau D_\tau(L) = 0.$$

Finally, (10) for $G = L$ reads

$$(53)\qquad 1 + \sum A_\rho(L) + C(L) = 0 \qquad (\rho \neq 1),$$

on account of (47). The last two equations give

$$(54)\qquad \sum x_\tau D_\tau(L) = n$$

and, on combining (51), (53), (54), we obtain

$$\sum_{\zeta} \zeta(T)\zeta(L) = q - 1 + n + 1 + (q-1)n = q(n+1).$$

Substitute for T the value (49), and use again (52). This gives

$$\sum_{\zeta} \zeta(T)\zeta(L) = \sum_{\zeta}\sum_{j=1}^{q-1} \zeta(Q^j)\zeta(L) = \sum_{j} n(Q^j)\delta(Q^j, L).$$

Hence

$$(55)\qquad \sum_{j=1}^{q-1} n(Q^j)\delta(Q^j, L) = q(n+1).$$

The equation (55) shows that not all $\delta(Q^j, L)$ can vanish. Hence L is conjugate to some power of Q

$$(56)\qquad L \sim Q^\nu \qquad (1 \leqq \nu \leqq q-1).$$

Since L is any element whose order is divisible by at least one prime factor of $p - 1 = qt$, and since Q has the order q, this proves that every prime factor of t divides q. Moreover, the formula (55) can be applied for $L = Q^i$, $i \not\equiv 0 \pmod q$. The left hand side is obviously the order of the normalizer of the cyclic group $\{Q^i\}$, and if this order is denoted by $N\{Q^i\}$, we have

$$N\{Q^i\} = q(n + 1). \tag{57}$$

As the order of a subgroup, $q(n + 1) = qst$ (cf. (42)) must divide the order $g = qtpr$ (cf. (45)) of $\mathfrak{G}$. This gives $s \mid pr$. If $p \mid s$, (42) and (41) would imply that $\beta \geqq p$, and this contradicts (40). Hence $s \mid r$. Now (44) shows that s divides q and is, therefore, a common divisor of r and $qt = p - 1$. By (46), we have $s = 1$.

Now (42), (44), and (45) become

$$n = t - 1, \qquad r = p - q, \qquad g = (p - 1)p(p - q) \tag{58}$$

while (40) and (41) yield

$$\alpha t - t + 2 = q = (p - 1)/t. \tag{59}$$

As we have seen, every prime divisor of t must divide q. Because of (59), t must be a power of 2, say

$$t = 2^{\mu-1}. \tag{60}$$

It follows from (38) that $\mu \geqq 2$.

Consider first the case $\mu = 2$, i.e. $t = 2$. Here, $n = 1$, $q = (p - 1)/2$ and $g = p(p - 1)(p + 1)/2$. From (59) we obtain $\alpha = (p - 1)/4$. But then (40) shows that $\beta = (p - 1)/4$. We have therefore, one character of degree 1, $(p - 5)/4$ characters of degree $p + 1$, $(p - 1)/4$ characters of degree $p - 1$, two characters of degree $(p + 1)/2$, and one character of degree p. The last fact follows from (15) which now reads

$$\sum x_\tau^2 = \frac{p}{2} - \frac{p-5}{4} - \frac{p-1}{4} - \frac{1}{2} = 1.$$

The method applied in an earlier paper[17] now gives $\mathfrak{G} \cong LF(2, p)$.

Assume now $\mu > 2$, i.e. $t \geqq 4$. As shown in (56), all elements L of an order 2^λ are conjugate to a power of Q. But (59) shows for $t \equiv 0 \pmod 4$ that $q \not\equiv 0 \pmod 4$. Hence no elements of order 4 exist, and a 2-Sylow-subgroup $\mathfrak{L}$ of $\mathfrak{G}$ can contain only elements of order 1 and 2. This implies that $\mathfrak{L}$ is an abelian group of type $(2, 2, \cdots, 2)$.

Let $\mathfrak{T}_i$ be the normalizer of the group $\{Q^i\}$, $i \not\equiv 0 \pmod q$. Every $\mathfrak{T}_i$ is contained in $\mathfrak{T} = \mathfrak{T}_1$, but since (57) shows that all $\mathfrak{T}_i$ have the same order $q(n + 1) = qt = p - 1$, we have

$$\mathfrak{T} = \mathfrak{T}_1 = \cdots = \mathfrak{T}_{q-1}.$$

---

[17] [2], sections 9 and 10.

Assume now that $q \neq 2$. Since we have $q \not\equiv 0 \pmod 4$, the number q will contain an odd prime divisor l. We set $L_0 = Q^{q/l}$. If X is an element of order 2 in $\mathfrak{T}$, we have

$$X^{-1}L_0X = L_0^j$$

for some j. But X^2 will commute with L_0 which gives $j^2 \equiv 1 \pmod l$, i.e. $j \equiv \pm 1 \pmod l$.

The order $p - 1 = qt$ of $\mathfrak{T}$ is divisible by 8. Because of the structure of the 2-Sylow subgroup $\mathfrak{L}$ of $\mathfrak{G}$, it follows that $\mathfrak{T}$ contains an abelian subgroup of type $(2, 2, 2)$. This implies that we have at least three elements X of order 2 in $\mathfrak{T}$ for which $j \equiv 1 \pmod l$.

If $j \equiv 1 \pmod l$, X and L_0 commute and $L = XL_0$ is an element of order $2l$. By (56), XL_0 then is conjugate in $\mathfrak{G}$ to a power of Q, say

$$(61) \qquad T^{-1}XL_0T^{-1} = Q^{\rho}.$$

Hence $T^{-1}L_0^2T = Q^{2\rho}$, i.e. $T^{-1}Q^{2q/l}T = Q^{2\rho}$. This shows that T belongs to the normalizer of $Q^{2q/l} \neq 1$, and therefore T belongs to $\mathfrak{T}$. Moreover, (61) implies that $T^{-1}X^lT = Q^{\rho l}$. Hence $Q^{\rho l}$ has order 2, i.e. $Q^{\rho l}$ must be equal to the only power $Q^{q/2}$ of Q of order 2. Since T transforms $Q^{q/2}$ into itself, we have $X = X^l = T^{-1}Q^{q/2}T = Q^{q/2}$. This gives a contradiction because we had shown that we have at least three different elements X of order 2 in $\mathfrak{T}$ for which $j \equiv 1$. It follows that $q = 2$ and then (58) and (60) give

$$(62) \quad p = qt + 1 = 2^{\mu} + 1, \qquad g = p(p-1)(p-2), \qquad t = \frac{p-1}{2}, \qquad n = \frac{p-3}{2}.$$

In particular, p must be a Fermat prime. Furthermore, (59) and (40) give

$$\alpha = 1, \qquad \beta = 1.$$

Hence $\mathfrak{G}$ has one representation A_1 of degree 1, one representation of degree $p - 1$, and $(p - 1)/2$ representations of degree $(1 + np)/t = p - 2$. The degree of all other irreducible representations is divisible by p.

The 2-Sylow subgroup $\mathfrak{L}$ must have the order $p - 1 = 2^{\mu}$. We may assume that $\mathfrak{L}$ contains Q. From (56) it follows that each of the $2^{\mu} - 1$ elements $L \neq 1$ of $\mathfrak{L}$ is conjugate to Q in $\mathfrak{G}$. Then L must be conjugate to Q in the normalizer $\mathfrak{N}(\mathfrak{L})$ of the abelian group $\mathfrak{L}$. Conversely, every element of $\mathfrak{N}(\mathfrak{L})$ transforms Q into an element $L \neq 1$ of $\mathfrak{L}$. But (57) for $i = 1$ and (62) give

$$N(Q) = q(n + 1) = 2(p - 1)/2 = p - 1 = 2^{\mu}.$$

Hence only the elements of $\mathfrak{L}$ will commute with Q. This shows that $\mathfrak{L}$ has the index $2^{\mu} - 1$ in $\mathfrak{N}(\mathfrak{L})$. Consequently, $\mathfrak{N}(\mathfrak{L})$ has the order $2^{\mu}(2^{\mu} - 1) = (p - 1)(p - 2)$ i.e. $\mathfrak{N}(\mathfrak{L})$ has the index p in $\mathfrak{G}$.

It now follows that $\mathfrak{G}$ possesses a permutation representation of degree p. If Π is the corresponding character, then Π contains A_1 exactly once. Since $p > 3$, Π cannot contain a character $C^{(\nu)}$ of degree $p - 2$, and it cannot contain any character D_{τ}. Hence we have

$$\Pi(G) = A_1(G) + B_1(G) = 1 + B_1(G).$$

From (8), (10), (47), and (48), we obtain $B_1(P^i) = -1$ for $i \not\equiv 0 \pmod p$, $B_1(L) = 0$, $B_1(M) = A_1(M) + C(M) = 1$.

Hence

(63) $\Pi(1) = p, \quad \Pi(P^i) = 0$ for $i \not\equiv 0 \pmod p, \quad \Pi(L) = 1, \quad \Pi(M) = 2$

where L and M have the same significance as in (47), (48). However, $\Pi(G)$ equals the number of letters not altered by the permutation representing G. Then (63) shows that we have a (1–1)-representation since $\Pi(G) = p$ only for $G = 1$. The subgroup leaving three letters fixed has the order 1, i.e. its index in $\mathfrak{G}$ is $p(p-1)(p-2)$. This implies that $\mathfrak{G}$ is three times transitive. From a theorem of Zassenhaus,[18] it now follows that $\mathfrak{G} \cong LF(2, p-1)$ and this finishes the proof of theorem 10.

6. Simple groups of order $(p-1)\,p\,(1+mp)/\tau$

We now drop the assumption (*) and propose to prove the following theorem

THEOREM 11. *Let $\mathfrak{G}$ be any non-cyclic simple group of order*

$$g = (p-1)p(1+mp)/\tau$$

where p is a prime, and where τ and m are any non-negative integers such that τ divides $p-1$. If m does not possess a representation of the form $m = (puh + u^2 + u + h)/(u+1)$ with positive integers u, h, in particular if $m < (p+3)/2$, then $\mathfrak{G}$ is either isomorphic to $LF(2, p)$ or to $LF(2, p-1)$, and in the second case p must be a Fermat prime, $p = 2^\mu + 1$, $(p > 3)$.

PROOF. Let $1 + np$ be the number of conjugate subgroups of order p in $\mathfrak{G}$. If $\mathfrak{P} = \{P\}$ is one of these Sylow-subgroups, then $g/(1+np)$ is the order of the normalizer of $\mathfrak{N}(\mathfrak{P})$. If the order of the normalizer (centralizer) of P is v, and if t classes of conjugate elements of $\mathfrak{G}$ contain elements of order p, we have

(64) $$g = (p-1)p(1+mp)/\tau = (p-1)pv(1+np)/t.$$

The number n here is positive since otherwise $\mathfrak{P}$ would be a normal subgroup of $\mathfrak{G}$. The number $1 + np$ of conjugate Sylow subgroups divides g and hence it divides $g\tau = (p-1)p(1+mp)$. Then lemma 1 shows that there exists an integer $h \geqq 0$ such that

(65) $$m = F(p, n, h).$$

Since it was assumed that m did not have a representation $m = F(p, u, h)$ with positive integers u and h, we must have $h = 0$ in (65). Then n becomes equal to m, and (64) gives

(66) $$n = m, \qquad t = v\tau.$$

Consider now the irreducible representations of $\mathfrak{G}$ which belong to the first

[18] H. Zassenhaus, Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität, vol. 11, p. 17 (1935).

p-block. For this block we have corresponding results[19] as for the A_ρ, B_σ, $C^{(\nu)}$ in section 1. There exists, however, other representations of a degree prime to p when $v > 1$, and (12) and (15) will no longer hold.

The degrees $a_\rho = pu_\rho + 1$ and $b_\sigma = pv_\sigma - 1$ in the first p-block must divide $g\tau = (p-1)p(1+pm)$. Then lemmas 1 and 2 show that we have $a_\rho = 1$ or $a_\rho = 1 + pm$ and $b_\sigma = p - 1$, cf. (30). Since $\mathfrak{G}$ is simple, only one a_ρ has the value 1, say $a_1 = 1$.

If $v = 1$, the group $\mathfrak{G}$ satisfies the assumption (*), and theorem 10 implies theorem 11 in this case. We therefore may assume that $v > 1$. Suppose now that $V \neq 1$ is a p-regular element of $\mathfrak{N}(P)$. If $\beta \neq 0$, it follows easily from (8, II) that $\mathfrak{B}_\sigma(P)$ has the characteristic roots $\epsilon, \epsilon^2, \cdots, \epsilon^{p-1}$ where ϵ is a primitive pth root of unity. We may arrange the characteristic roots $x_1, x_2, \cdots, x_{p-1}$ of $\mathfrak{B}_\sigma(V)$ in such a manner that P^iV has the characteristic roots $x_\mu\epsilon^{i\mu}$, ($\mu = 1, 2, \cdots, p-1$). However, $\mathfrak{B}_\sigma(P^iV) = -1$ for $i \not\equiv 0 \pmod p$.[20] This gives

$$(67) \qquad \sum_{\mu=1}^{p-1} \epsilon^{\mu i} x_\mu = -1 = \sum_{\mu=1}^{p-1} \epsilon^{\mu i} \qquad (i = 1, 2, \cdots, p-1).$$

The x_μ are themselves roots of unity, of an exponent prime to p. Since $\epsilon, \epsilon^2, \cdots, \epsilon^{p-1}$ must be linearly independent over the field generated by the x_μ, (67) implies $x_\mu = 1$. But then $\mathfrak{B}_\sigma(V) = I$, and since $\mathfrak{G}$ was simple, we have $V = 1$ which gives a contradiction. Hence we have $\beta = 0$.

It now follows from (9) that

$$a_1 = 1, \qquad a_2 = \cdots = a_q = pm + 1;$$

and (11) gives $\delta = -1$ and

$$(68) \qquad c = 1 + (q-1)(pm+1).$$

This shows that $1 + pm$ and c are relatively prime. The degree c divides g, and (64) now implies that c divides $(p-1)/\tau$. However, since $c \neq 1$ and $n = m \neq 0$, the equation (68) yields $c > p > (p-1)/\tau$. We have a contradiction, and theorem 11 is proved.

7. Examples of groups which satisfy the condition (*)

First, let $\mathfrak{G}$ be a transitive permutation group on p letters, where p is a prime. The order g of $\mathfrak{G}$ then is divisible by p, and $\mathfrak{G}$ contains elements P of order p. Each such element P is represented by a simple cycle of length p. It now follows easily that P commutes only with its own powers; i.e. $\mathfrak{G}$ satisfies the condition (*). In a similar manner, we can show that a doubly transitive group of degree $p - 1$ satisfies the condition (*).

We next consider irreducible groups $\mathfrak{G}$ of linear transformations of a p-dimensional vector space[21] where p again is a prime. Assume that $\mathfrak{G}$ is of finite

---

[19] cf. [1], theorems 4 and 11.
[20] cf. [1], theorems 4 and 11.
[21] cf. footnote 2.

order g and that its center consists of the unit element only. The order g is divisible by the degree p of the irreducible group. Let $\mathfrak{P}$ be a Sylow-subgroup of $\mathfrak{G}$, and let P_0 be an invariant element of $\mathfrak{P}$ which is different from 1. Then P_0 cannot be a scalar multiple of the unit matrix, since it does not belong to the center $\{1\}$ of $\mathfrak{G}$. But P_0 commutes with every element of $\mathfrak{P}$, and Schur's lemma implies that the linear group $\mathfrak{P}$ is reducible. The degree of every irreducible constituent of $\mathfrak{P}$ must be 1, since it divides the order of $\mathfrak{P}$. Hence $\mathfrak{P}$ can be taken as a set of diagonal matrices, i.e. $\mathfrak{P}$ is an abelian group. We now prove

LEMMA 5.[22] *Let $\mathfrak{G}$ be a group of order $g = p^a g^*$ with $(p, g^*) = 1$, and assume that $\mathfrak{G}$ does not contain invariant elements of order p, and that the Sylow-subgroup $\mathfrak{P}$ of order p^a in $\mathfrak{G}$ is abelian. If $\mathfrak{Z}$ is an irreducible* (1–1) *representation of $\mathfrak{G}$ of degree p^μ, then $\mu = a$.*

PROOF. Let ζ be the character of $\mathfrak{Z}$. If G is an element of $\mathfrak{G}$ which has exactly j conjugate elements then it is well known that $j\zeta(G)/\zeta(1) = j\zeta(G)/p^\mu$ is an algebraic integer. Since $\mathfrak{P}$ is abelian, the number j is prime to p when G lies in $\mathfrak{P}$. Hence $\zeta(G) \equiv 0 \pmod{p^\mu}$ for G in $\mathfrak{P}$. On the other hand, $\zeta(G)$ is a sum of p^μ roots of unity. If $G \neq 1$, not all these roots of unity can be equal. Hence the integer $\zeta(G)/p^\mu$ and all its algebraic conjugates are smaller than 1 in absolute value which implies $\zeta(G) = 0$ for every $G \neq 1$ in $\mathfrak{P}$.[23] Then the character ζ is of the highest kind,[24] i.e. $p^\mu = \zeta(1) \equiv 0 \pmod{p^a}$ which yield $\mu = a$, as was stated.

LEMMA 6. *Let $\mathfrak{G}$ be a group of order $g = p^a g^*$ with $(p, g^*) = 1$. If the center of $\mathfrak{G}$ consists of the unit element only, and if $\mathfrak{G}$ has a* (1–1)*-representation $\mathfrak{Z}$ of degree p^a, then the centralizer of a Sylow subgroup $\mathfrak{P}$ of order p^a is contained in $\mathfrak{P}$.*

PROOF. The character ζ of $\mathfrak{Z}$ has the values[25]

$$\zeta(P) = \begin{cases} p^a & P = 1 \\ 0 & P \text{ in } \mathfrak{P}, \quad P \neq 1. \end{cases}$$

Hence $\mathfrak{Z}(P)$ is the regular representation of $\mathfrak{P}$; we may assume that it breaks up into the distinct irreducible representations $\mathfrak{F}_\mu$ of $\mathfrak{P}$, each $\mathfrak{F}_\mu$ appearing f_μ times where f_μ is the degree of $\mathfrak{F}_\mu$. We then have

$$\mathfrak{Z}(P) = \begin{pmatrix} \ddots & & 0 \\ & f_\mu \times F_\mu & \\ 0 & & \ddots \end{pmatrix}, \qquad F_\mu = \mathfrak{F}_\mu(P).$$

[22] The following lemmas 5 and 6 are proved here in a more general form than necessary for our purpose. However, in the form given here, they can also be used in other connections.

[23] For this argument, cf. W. Burnside, Proceedings of the London Mathematical Society (2) vol. 1, p. 388–392 (1904).

[24] cf. theorem 10 of the paper mentioned in footnote 16.

[25] cf. footnote 16.

Let V be an element of the centralizer $\mathfrak{C}$ of $\mathfrak{P}$ such that the order v of V is prime to p. Then $\mathfrak{Z}(V)$ will commute with $\mathfrak{Z}(P)$. It follows that $\mathfrak{Z}(V)$ is of the form

$$\mathfrak{Z}(V) = \begin{pmatrix} \ddots & & 0 \\ & T_\mu \times I_\mu & \\ 0 & & \ddots \end{pmatrix}$$

where T_μ is a matrix of degree f_μ and I_μ is the unit matrix of degree f_μ. Then $\mathfrak{Z}(V^iP)$ breaks up completely into the matrices $T_\mu^i \times F_\mu$. If $\tau_\mu^{(i)}$ is the trace of T_μ^i and if $\theta_\mu(P)$ is the character of $F_\mu(P)$, we find

$$\zeta(V^iP) = \sum \tau_\mu^{(i)}\theta_\mu(P).$$

Since ζ is of the highest kind, we have

$$\sum_\mu \tau_\mu^{(i)}\Theta_\mu(P) = 0 \qquad \text{for } P \text{ in } \mathfrak{P}, \quad P \neq 1. \tag{70}$$

Set $\sum \tau_\mu^{(i)}f_\mu = \tau^{(i)}$ where the sum extends over all values of μ. Using the orthogonality relations for the characters of $\mathfrak{P}$, we derive from (70) the equation

$$p^a\tau_\nu^{(i)} = \sum_P \sum_\mu \tau_\mu^{(i)}\Theta_\mu(P)\Theta_\nu(P^{-1}) = \sum_\mu \tau_\mu^{(i)}f_\mu f_\nu = f_\nu\tau^{(i)}.$$

This shows that the matrices T_ν^i/f_ν have the same trace for all values of ν. If $\mathfrak{F}_1$ is the 1-representation of $\mathfrak{P}$, then T_1 is a vth root of unity λ where v is the order of V. Hence

$$\mathrm{tr}(T_\nu^i) = \tau_\nu^{(i)} = f_\nu\tau^{(i)}/p^a = f_\nu\tau_1^{(i)} = f_\nu\lambda^i.$$

The mapping $V^i \to T_\nu^i$ defines a representation of the group $\{V\}$. Since its character is identical with the character of the representation $V^i \to \lambda^i I_\nu$, it follows easily that $T_\nu^i = \lambda^i I_\nu$. Hence $\mathfrak{Z}(V) = \lambda I$. This is impossible for $V \neq 1$, because $\mathfrak{Z}$ was a (1–1)-representation, and $\mathfrak{G}$ did not contain any invariant elements except 1.

Hence the centralizer $\mathfrak{C}$ of $\mathfrak{P}$ cannot contain elements of an order prime to p. Consequently, the order of $\mathfrak{C}$ itself is a power of p. Since $\mathfrak{C}$ and $\mathfrak{P}$ generate a p-group contained in $\mathfrak{G}$, we have $\mathfrak{C} \subseteq \mathfrak{P}$ and this proves lemma 6.

Returning to irreducible linear groups $\mathfrak{G}$ of degree p whose center consists only of the unit element, it follows from lemma 5 that the order g contains p to the first power only. If P is an element of order p, then lemma 6 shows that $\{P\}$ is the centralizer of P. Hence we see

THEOREM 12. *If p is a prime, then the transitive permutation groups p, the doubly transitive permutation groups of degree $p + 1$, and the irreducible linear groups in p dimensions with the center $\{1\}$ satisfy the condition (*).*

THE INSTITUTE FOR ADVANCED STUDY.

ANNALS OF MATHEMATICS
Vol. 44, No. 1, January, 1943

ON THE CONVERGENCE OF CONTINUED FRACTIONS TO MEROMORPHIC FUNCTIONS*

BY WALTER LEIGHTON AND W. J. THRON

(Received August 28, 1942)

1. Introduction

In this paper we give a new set of convergence criteria for continued fractions of the form

$$\frac{1}{1} + \frac{a_2}{1} + \frac{a_3}{1} + \cdots, \tag{1}$$

where the a_n are complex numbers. The fundamental results are contained in Theorem 3. Application is made to continued fractions of the form

$$1 + \frac{a_1 x}{1} + \frac{a_2 x}{1} + \cdots, \tag{2}$$

where x is a complex variable. The basic algorithm in §3 is equivalent to one introduced by Euler [1, p. 206]. A direct proof is included since it is very brief.

2. General concepts

The continued fraction

$$b_0 + \frac{a_1}{b_1} + \frac{a_2}{b_2} + \cdots,$$

where the a_n and b_n are complex numbers, is said to converge or diverge according as the series

$$\frac{A_0}{B_0} + \left(\frac{A_1}{B_1} - \frac{A_0}{B_0}\right) + \left(\frac{A_2}{B_2} - \frac{A_1}{B_1}\right) + \cdots \tag{3}$$

converges or diverges. If the series converges to a value v, the continued fraction is said to converge to v. The numbers A_n and B_n are defined by the recursion relations

$$\begin{aligned} A_0 &= b_0\,, A_1 = b_0 b_1 + a_1\,, \\ B_0 &= 1, B_1 = b_1\,, \\ A_n &= b_n A_{n-1} + a_n A_{n-2}\,, \\ B_n &= b_n B_{n-1} + a_n B_{n-2}\,. \end{aligned} \tag{4}$$

The ratio A_n/B_n is called the n[th] approximant of the continued fraction. Finally, if it is known that the series either converges or, at worst, diverges to ∞, the

* Presented to The American Mathematical Society, Sept. 8, 1942.

continued fraction is said to *converge at least in the wider sense*. If the numbers a_n and b_n are functions of a variable x, and if for x in a prescribed set the series (3) converges uniformly with respect to x, the continued fraction is said to converge uniformly with respect to x in this set.

3. An algorithm

It is well known that the quantity

$$D_n = \frac{A_n}{B_n} - \frac{A_{n-1}}{B_{n-1}} \qquad (n = 1, 2, 3, \cdots),$$

is given by the relation

$$D_n = \frac{(-1)^{n-1} a_1 a_2 \cdots a_n}{B_{n-1} B_n}, \tag{5}$$

at least when $B_{n-1}B_n \neq 0$. The numbers D_n assume a particularly simple form for the continued fraction

$$\frac{1}{1} + \frac{z_2 - 1}{1} + \frac{z_2(z_3 - 1)}{1} + \frac{z_3(z_4 - 1)}{1} + \cdots. \tag{6}$$

LEMMA. *For the continued fraction* (6) *the quantity D_n is given by the formula*

$$D_n = (-1)^{n-1} \frac{(z_2 - 1)(z_3 - 1) \cdots (z_n - 1)}{z_2 z_3 \cdots z_n} \qquad (n \geqq 2).$$

This result will follow from (5) once it has been shown that

$$B_n = z_2 z_3 \cdots z_n \qquad (n \geqq 2). \tag{5)'$$

The proof of (5)′ is by induction. This formula may be verified readily for $n = 2$ and for $n = 3$. Suppose it holds for $n = k - 1$ and for $n = k$. By the last relation (4)

$$B_{k+1} = B_k + z_k(z_{k+1} - 1)B_{k-1}.$$

The proof can now be completed by applying the induction hypothesis to B_k and to B_{k-1}.

It is an immediate consequence of the lemma that for the continued fraction (6) the series (3) takes the form

$$\sum_{n=1}^{\infty} D_n = 1 + \sum_{n=2}^{\infty} (-1)^{n-1}\left(1 - \frac{1}{z_2}\right)\left(1 - \frac{1}{z_3}\right) \cdots \left(1 - \frac{1}{z_n}\right).$$

THEOREM 1. *The continued fraction* (6) *converges if the numbers z_n $(n \geqq 2)$ satisfy the relation*

$$\left| z_{2n} + \frac{1}{h^2 - 1} \right| > \frac{h}{h^2 - 1} + e \tag{7}$$

and

$$\left| z_{2n+1} - \frac{h^2}{h^2 - 1} \right| < \frac{h}{h^2 - 1} - d, \tag{8}$$

where h is any number > 1, and e and d are arbitrary positive numbers.

If relations (7) and (8) hold, positive numbers e' and d' can be found such that

$$\left| 1 - \frac{1}{z_{2n}} \right| < h - e'$$

and

$$\left| 1 - \frac{1}{z_{2n+1}} \right| < \frac{1}{h + d'}.$$

Under these conditions

$$| D_n | < \frac{(h - e')^{[n/2]}}{(h + d')^{[(n-1)/2]}}.$$

The series (3) is therefore majorized by the convergent series

$$1 + (h - e') + \frac{(h - e')}{(h + d')} + \frac{(h - e')^2}{(h + d')} + \cdots; \tag{9}$$

hence, the series (3) and the continued fraction (6) converge. The proof is complete.

THEOREM 2. *If the quantities z_n are functions of a complex variable x and if for all values of x contained in a certain region D, the functions $z_{2n}(x)$ and $z_{2n+1}(x)$ satisfy relations (7) and (8) respectively, where h, e and d are independent of x, the continued fraction (6) converges uniformly for all x in D. Further, if the functions $z_n(x)$ are analytic for all x in D, the continued fraction (7) converges to a function that is analytic for all x in D.*

Under the conditions of the theorem, the series (9) is independent of x and majorizes the series (3) for all x in D. Thus the series (3) converges uniformly. The partial sums of the series (3) are rational functions of the quantities z_n and hence analytic functions of x. Weierstrass' well-known theorem then insures the validity of the second part of the theorem.

4. Convergence criteria

The results of the preceding section will now be applied to continued fractions (1). It is our aim to find conditions on the numbers a_n such that for any given sequence $\{a_n\}$ satisfying these conditions a sequence $\{z_n\}$ is uniquely determined by the system of equations

$$\begin{aligned} a_2 &= z_2 - 1, \\ a_{2n} &= z_{2n-1}(z_{2n} - 1), \\ a_{2n+1} &= z_{2n}(z_{2n+1} - 1), \end{aligned} \tag{10}$$

in such a way that the numbers z_n satisfy the conditions of Theorem 1. Conditions of this kind on the numbers a_n will clearly entail convergence of the continued fraction (1).

We proceed with the following definitions:

$$
\text{(11)} \qquad
\begin{aligned}
&\text{(i)} \quad z \in Z_1, \quad \text{if} \quad \left| z - \frac{h^2}{h^2 - 1} \right| < \frac{h}{h^2 - 1}, \\
&\text{(ii)} \quad z \in Z_2, \quad \text{if} \quad \left| z + \frac{1}{h^2 - 1} \right| > \frac{h}{h^2 - 1}, \\
&\text{(iii)} \quad w \in W_1, \quad \text{if} \quad \left| w - \frac{1}{h^2 - 1} \right| < \frac{h}{h^2 - 1}, \\
&\text{(iv)} \quad w \in W_2, \quad \text{if} \quad \left| w + \frac{h^2}{h^2 - 1} \right| > \frac{h}{h^2 - 1}.
\end{aligned}
$$

It follows that $z \in Z_1$ implies $z - 1 \in W_1$ and that $z \in Z_2$ implies $z - 1 \in W_2$, and conversely.

We shall determine regions $A_1(h)$ and $A_2(h)$ with the property that if $a_{2n+1} \in A_1(h)$ and $a_{2n} \in A_2(h)$, the numbers $z_n(\neq 0)$ defined by equations (10) will belong to Z_1, when n is odd, and to Z_2, when n is even.

To this end let us denote by $z \cdot W$ the set of points containing all products zw, where z is fixed and w ranges over the set W. By $D[z \cdot W_2]$ we shall denote the point set intersection of all such sets $z \cdot W_2$ as z is allowed to vary over the region Z_1. It follows that a is an element of $D[z \cdot W_2]$ if and only if a/z is an element of W_2 for all $z \in Z_1$.

The region $A_2(h)$ may be taken as $D[z \cdot W_2]$, since for any $a \in A_2(h)$ and $z \in Z_1$ there exists a $z' \in Z_2$ (i.e., $z' - 1 \in W_2$) such that $a = z(z' - 1)$.

Similarly the region $A_1(h)$ may be taken as $D[z \cdot W_1]$ where z assumes all values of the set X_2. As $z = 1 \in Z_1$ it is clear that the set $A_2(h)$ is contained in the set W_2.

If now a sequence $\{a_n\}$ with $a_{2n} \in A_2(h)$ and $a_{2n+1} \in A_1(h)$ is given, a sequence $\{z_n\}$ can be uniquely determined ($z_n \neq 0$) satisfying the system of equations (10). If further the numbers a_{2n} and a_{2n+1} are bounded away from the boundaries of their respective regions, then an e and a d can be found such that the corresponding z_n satisfy conditions (7) and (8) respectively.

From the definition of the set $A_2(h)$ it follows that it is that part of the complex plane which remains when all circular regions $z \cdot C$ are deleted, where z is allowed to vary over the closure of the region Z_1 and C is the closed circular region

$$\left| z + \frac{h^2}{h^2 - 1} \right| \leqq \frac{h}{h^2 - 1}.$$

The boundary of the region $A_2(h)$ can be determined as follows. Consider the products zc where z and c are arbitrary values from the regions Z_1 and C

respectively. These products lie outside $A_2(h)$ and can lie on the boundary of $A_2(h)$ only if both z and c lie on the boundaries of their respective regions. Let us fix the sum of the arguments of z and c. The products then lie on a ray passing through the origin, and only those products will lie on the boundary of $A_2(h)$ for which the absolute value of the product zc, for a fixed arg zc, is maximized and minimized respectively.

We now note that the circles that form the boundaries of Z_1 and C have the same radius and have centers on the real axis with the x coordinates $h^2/(h^2 - 1)$ and $-h^2/(h^2 - 1)$ respectively.

We further note that for any fixed arg z there are two points z on the circle bounding Z_1. A similar remark is true for C. Let $| z_1 | < | z_2 |$, arg z_1 = arg z_2 and $| c_1 | < | c_2 |$, arg c_1 = arg c_2. It is clear that the minimum and maximum

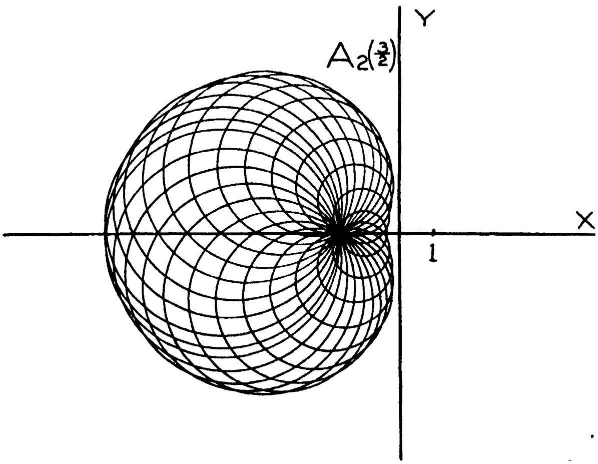


FIG. 1. The Region $A_2(\frac{3}{2})$

absolute values are obtained respectively by considering products of the type z_1c_1 and z_2c_2.

Let us set $k = h(h^2 - 1)$ and $a = h^2/(h^2 - 1)$. Then $z_1 = d(\alpha)e^{i\alpha}$ and $c_1 = d(\beta)e^{i(\beta+\pi)}$, where

$$d(\theta) = a \cos \theta - (k^2 - a^2 \sin^2 \theta)^{\frac{1}{2}}.$$

It follows from considerations of elementary calculus that for $\alpha + \beta = \gamma$ fixed, the expression $| z_1c_1 |$ is a minimum when $\alpha = \beta = \gamma/2$. We note that $\gamma/2$ cannot exceed the value arc sin $1/h$.

Substituting the values $\alpha = \beta = \gamma/2 = (\theta - \pi)/2$ and $a = h^2/(h^2 - 1)$, $k = h/(h^2 - 1)$ we then have as part of the boundary of the region $A_2(h)$

$$r = d^2(\gamma/2) = \left(\frac{h}{h^2 - 1}\right)^2 [1 - h^2 \cos \theta - 2h \sin \theta/2(1 - h^2 \cos^2 \theta/2)^{1/2}]$$

where $\pi - 2\lambda \leqq \theta \leqq \pi + 2\lambda$; $\lambda = \text{arc sin } 1/h$.

An analogous argument shows that the remaining portion of the boundary is given by the equation

$$r = \left(\frac{h}{h^2 - 1}\right)^2 [1 - h^2 \cos\theta + 2h \sin\theta/2(1 - h^2 \cos^2\theta/2)^{1/2}],$$

$$(\pi - 2\lambda \leqq \theta \leqq \pi + 2\lambda).$$

It follows that $A_2(h)$ is the region outside the curve

$$r^2 + 2r(a^2 \cos\theta - k^2) + a^2 = 0, \tag{12}$$

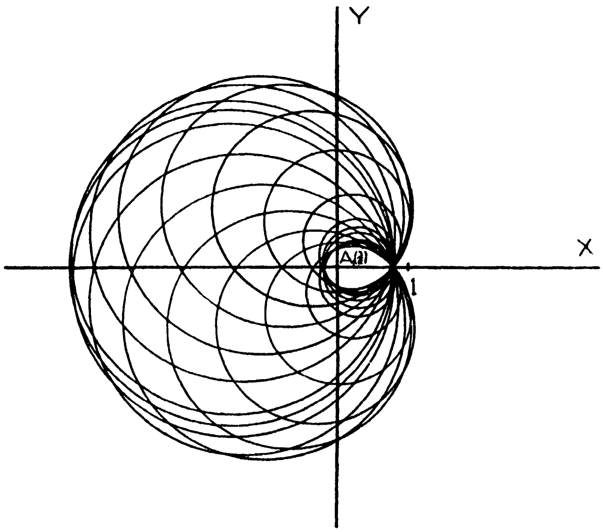


FIG. 2. The Region $A_1(\frac{3}{2})$

where we recall that

$$a = \frac{h^2}{h^2 - 1}, \qquad k = \frac{a}{h},$$

and h is any number > 1.

Similarly, the condition that $a_{2n+1} = re^{i\theta}$ be an element of the region $A_1(h)$ can be shown to be

$$r < \frac{1}{(h^2 - 1)^2}[h^2 - \cos\theta - 2\sin\theta/2(h^2 - \cos^2\theta/2)^{1/2}]. \tag{13}$$

The following theorem has now been proved.

THEOREM 3. *The continued fraction*

$$\frac{1}{1} + \frac{a_2}{1} + \frac{a_3}{1} + \cdots$$

converges if the numbers $a_{2n} \in A_2(h)$ [*formula* (12)] *and the numbers* $a_{2n+1} \in A_1(h)$ [*formula* (13)], *for some* $h > 1$, *and if in addition the numbers* a_n *are bounded away from the boundaries of these regions.*

It is of course evident that the roles of the odd and of the even a's can be interchanged.

As $h \to 1$ the regions $A_2(h)$ and $A_1(h)$ both tend to the parabola found by Scott and Wall [2].

We now note that if the parameter h is taken large enough, the region $A_2(h)$ can be made to include all the plane except an arbitrary small neighborhood of the point $z = -1$. The region $A_1(h)$ is then restricted to a small neighborhood of the point $z = 0$. The following is a corollary of Theorem 3.

COROLLARY. *The continued fraction* (1) *converges at least in the wider sense, if* $\lim a_{2n+1} = 0$ *and if* $z = -1$ *is not a limit point of the sequence* $\{a_{2n}\}$.

Under the conditions of the corollary there exists an $e > 0$ and a positive integer $n(e)$ such that $|1 + a_{2n}| > e$, when $n \geqq n(e)$. Further, there exists a number h such that $a_{2n} \in A_2(h)$ for all $n \geqq n(e)$. Finally we can find a positive integer $n(h)$ such that for all $n \geqq n(h)$, $a_{2n+1} \in A_1(3h)$. If we set $N = \max [n(e), n(h)]$, the numbers a_{2n} and $a_{2n+1} (n \geqq N)$ will certainly lie in the regions $A_2(2h)$ and $A_1(2h)$ respectively and will be bounded away from the boundaries of these regions. Theorem 3 therefore insures the convergence of the continued fraction

$$\frac{1}{1} + \frac{a_{2N}}{1} + \frac{a_{2N+1}}{1} + \cdots.$$

From a well-known argument it follows that the original continued fraction converges at least in the wider sense. The proof of the corollary is complete.

This result permits the following application to continued fractions of the form (2).

THEOREM 4. *Let* L (*possibly including the point* ∞) *be the set of limit points of the sequence* $\{a_{2n}\}$ *of the continued fraction*

$$1 + \frac{a_1 x}{1} + \frac{a_2 x}{1} + \cdots,$$

and suppose $\lim a_{2n+1} = 0$. *The continued fraction then converges to a meromorphic function of the complex variable* x *in every region contained in the set* D, *where* D *is defined as follows*: $x \in D$ *if*

$$|x| < M,$$

and if

$$\left| x + \frac{1}{k} \right| > e$$

for all $k \in L$, $k \neq 0$. *The constants* M *and* e *are any positive numbers such that* $e \leqq 4M$.

Consider the set of numbers a_{2n}. They may be distributed into three categories as follows. If the point at infinity belongs to L, the first category A will contain all a_{2n} such that $|a_{2n}| \geqq 2/e$. If infinity does not belong to L, A is to be the null set. By hypothesis $|x| > e$ for all x in D and hence $|1 + a_{2n}x| > 1 > e/4M$ for $a_{2n} \epsilon A$. The second category B contains all a_{2n} such that $|a_{2n}| < 1/2M$. This set may be null. Each member of B has the property that $|a_{2n}x| < 1/2$ and hence that $|1 + a_{2n}x| > 1/2 > e/4M$, for all x in D. Finally, the third category C contains the remaining numbers a_{2n}. Except for at most a finite number of elements of C, the members of C and hence the limit points of C are in modulus $\geqq 1/2M$ and $\leqq b$, where b is a suitably chosen finite constant $\geqq 2/e$.

By Zermelo's principle, if needed, numbers k_n belonging to L and a positive integer n_0 can be chosen such that for $n \geqq n_0$

$$\left| \frac{a_{2n}}{k_n} - 1 \right| < \frac{e}{4M} \qquad (a_{2n} \epsilon C).$$

It follows from our hypothesis on D that

$$\left| a_{2n}x + \frac{a_{2n}}{k_n} \right| > \frac{e}{2M}$$

for all x in D. Combining these two relations we see that

$$|1 + a_{2n}x| \geqq \left| a_{2n}x + \frac{a_{2n}}{k_n} \right| - \left| \frac{a_{2n}}{k_n} - 1 \right| > \frac{e}{4M},$$

for a_{2n} in $C(n \geqq n_0)$ and all x in D.

A number h independent of x thus exists such that $a_{2n}x \epsilon A_2(h)$, where for all x in D and all $n \geqq n_0$ the quantities $a_{2n}x$ are bounded away from the boundary of $A_2(h)$. Further, since $\lim a_{2n+1} = 0$ and $|x| < M$, a positive integer n_1 exists such that for all $n \geqq n_1$ and all x in D, $a_{2n+1}x \epsilon A_1(h)$ where the set of all $a_{2n+1}x$ is bounded away from the boundary of $A_1(h)$. If we now let $N = \max(n_0, n_1)$, it follows that the continued fraction

$$\frac{1}{1} + \frac{a_{2N}x}{1} + \frac{a_{2N+1}x}{1} + \cdots \tag{14}$$

converges. The uniform convergence of this continued fraction follows from Theorem 2. The quantities z mentioned there are rational functions of the variable x,

$$z_n = 1 + \frac{a_n x}{1} + \frac{a_{n-1}x}{1} + \cdots + \frac{a_2 x}{1},$$

and have no pole for x in D as for all x in D the system of equations (10) has unique and finite solutions. The continued fraction (14) therefore converges to a regular analytic function in every region contained in D. A well-known argument then insures the convergence of the complete continued fraction to a

function meromorphic in every region contained in D. This completes the proof of Theorem 4.

To illustrate what may happen when $\lim a_{2n} = -1$ and $\lim a_{2n+1} = 0$, we give two examples. The continued fraction (1) diverges if $a_2 = -2, a_{2n} = -(n+1)/(n-1)$ and $a_{2n+1} = -1/n^2$; it converges if $a_2 = -2$, $a_{2n} = -(n+1)(n^2 - 3n + 3)/(n-1)^3$ and $a_{2n+1} = -1/n^4$. These results can be quickly verified by a reference to equations (10).

5. Value regions

As a by-product of the method used to establish the above convergence criteria we obtain results on value regions. The connection is established by the following lemma.

LEMMA. *If there exist in the complex plane four regions A_1, A_2, V_1, V_2 such that*

$$\text{i)}\quad \begin{aligned}&(a)\ 1 + A_1 \subset V_1,\\ &(b)\ 1 + A_2 \subset V_2,\end{aligned}$$

$$\text{ii)}\quad \begin{aligned}&(a)\ 1 + a/v \in V_1,\ \text{if } a \in A_1,\ v \in V_2,\\ &(b)\ 1 + a/v \in V_2,\ \text{if } a \in A_2,\ v \in V_1,\end{aligned}$$

then all approximants of the continued fraction

$$1 + \frac{a_1}{1} + \frac{a_2}{1} + \cdots \tag{15}$$

lie in the region V_1, and all approximants of

$$1 + \frac{a_2}{1} + \frac{a_3}{1} + \cdots \tag{16}$$

lie in V_2, if $a_{2n} \in A_2$, $a_{2n+1} \in A_1$.

Let

$$\frac{A_{2n}}{B_{2n}} = 1 + \frac{a_1}{1} + \cdots + \frac{a_{2n}}{1}.$$

It follows from condition (b) of i) that $1 + a_{2n} \in V_2$. Alternate application of conditions ii) (a) and ii) (b) then gives the desired result. The proof is analogous for the odd approximants of the first and all approximants of the second continued fraction.

We now note that the conditions of the lemma are satisfied for

$$A_1(h) = A_1,\ A_2(h) = A_2,$$

$$Z_1 = V_1,\ Z_2 = V_2.$$

The following theorem is then an immediate consequence of our previous results and the lemma.

THEOREM 5. *If* $h > 1$, $a_{2n} \epsilon A_2(h)$ *[formula* (12)*]*, $a_{2n+1} \epsilon A_1(h)$ *[formula* (13)*], then all approximants of the continued fraction* (15) *lie in the region* Z_1 *[formula* (11), (i)*] and all approximants of* (16) *lie in* Z_2 *[formula* (11), (ii)*]*.

THE RICE INSTITUTE.

BIBLIOGRAPHY

1. O. PERRON, *Die Lehre von den Kettenbrüchen*, Leipzig, 1929.
2. W. T. SCOTT AND H. S. WALL, *A convergence theorem for continued fractions*, Trans. A. M. S. vol. 47 (1940) pp. 155–172.

ANNALS OF MATHEMATICS
Vol. 44, No. 1, January, 1943

ON THE CONTINUUM HYPOTHESIS

BY ISAIAH MAXIMOFF

(Received March 21, 1942)

In an earlier paper[1] the author proved that the set $E_x^{(r)}$ of all the sequences

(1) $$x = \{x_0, x_1, x_2, \cdots x_\alpha, \cdots\}, \alpha \leqq \beta < \Omega_r,$$

(2) $$x = \{x_0, x_1, x_2, \cdots x_\alpha, \cdots\}, \alpha < \Omega_r,$$

where x_α is any one of the numbers

(3) $$1, 2, 3, \cdots \gamma, \cdots, \gamma < \Omega_r,$$

is a *continuum* of power $2^{\aleph_i}$.

Let (a, b) be a pair of elements of the set $E_x^{(r)}$ such that $a < b$.

DEFINITION 1. *The set of all the elements x such that $a < x < b$ will be called an interval (a, b) with the boundaries a, b.*

Let E be a set of points in $E_x^{(r)}$.

DEFINITION 2. *A point x_0 of $E_x^{(r)}$ will be called a limit ($\mathfrak{I}$)-point for E if every interval containing x_0 contains at least one point x_1 of E different from x_0, $x_1 \neq x_0$.*

Let $E_{\mathfrak{I}}^{(1)}$ be the set of all the limit ($\mathfrak{I}$)-points for E.

DEFINITION 3. *We shall say that a set E is perfect ($\mathfrak{I}$) if $E_{\mathfrak{I}}^{(1)} = E$.*

Let $y = f(x)$ be a single valued function defined in $E \subset E_x^{(r)}$. Denote by $f(E)$ the set of all the points $y = f(x)$ where $x \in E$.

DEFINITION 4. *The function $y = f(x)$ will be called regular in E if for every pair (x_1, x_2) of points x_1, x_2 of E such that $x_1 \neq x_2$ we have $f(x_1) \neq f(x_2)$.*

DEFINITION 5. *The function $y = f(x)$ will be called continuous ($\mathfrak{I}$) in E if for every pair (x_0, y_0) where $y_0 = f(x_0)$, $x_0 \in E$, to every interval $i(y_0)$ containing y_0 there corresponds an interval $i(x_0)$ containing x_0 and having the following property: from $x_0 \in i(x_0)E$ there follows $y \in i(y_0)f(E)$. We then write: $i(y_0) \rightarrow i(x_0)$.*

DEFINITION 6. *The set $E \subset E_x^{(r)}$ will be called hypermeasurable ($\mathfrak{I}$) if E either contains a perfect ($\mathfrak{I}$) set or is of the power $\leqq \aleph_r$.*

DEFINITION 7. *A function $y = f(x)$ defined in E, will be called hypermeasurable ($\mathfrak{I}$) in E, if the set E contains a perfect ($\mathfrak{I}$) set $\mathfrak{P}$ in which $f(x)$ is continuous ($\mathfrak{I}$).*

DEFINITION 8. *Let*

(4) $$x = \{x_0, x_1, x_2, \cdots, x_\alpha, \cdots\}, \alpha < \Omega_{r+k},$$

be a sequence of points x_α of the space $E_x^{(r)}$. This sequence will be called a point of the transfinite space E_x^{rk}.

Evidently, the power of the set E_x^{rk} is equal to $2^{\aleph_{r+k}}$. We shall say that a point

(5) $$x' = \{x'_0, x'_1, x'_2, \cdots x'_\alpha \cdots\}, \alpha < \Omega_{r+k},$$

of E_x^{rk} is equal to a point

[1] *On a Continuum of Power $2^{\aleph_i}$* (Annals of Mathematics, vol. 41, no. 2, April 1940).

$$(6) \qquad x'' = \{x_0'', x_1'', x_2'', \cdots x_\alpha'', \cdots\}, \alpha < \Omega_{r+k},$$

if $x_\alpha' = x_\alpha''$ for every $\alpha < \Omega_{r+k}$. In the contrary case we write $x' \neq x''$. We shall denote by $\rho\, I$ the point

$$x = \{x_0, x_1, x_2, \cdots x_\alpha, \cdots\}, \alpha < \Omega_{r+k},$$

of E_x^{rk} such that $x_0 = x_1 = x_2 = \cdots = x_\alpha = \cdots = \rho \epsilon E_x^{(r)}$.

Let E be any set of points in E_x^{rk} and let

$$\bar{x} = \{\bar{x}_0, \bar{x}_1, \bar{x}_2, \cdots \bar{x}_\alpha, \cdots\}, \alpha < \Omega_{r+k},$$

be any point of E. We shall denote
(i) by $\{\bar{x}\}_\alpha$ the point $\bar{x}_\alpha$ of $E_x^{(r)}$, $\bar{x}_\alpha = \{\bar{x}\}_\alpha$;
(ii) by $\{E\}_\alpha$ the set of all the points $\{\bar{x}\}_\alpha$ where $\bar{x} \epsilon E$. A set $\mathcal{I} \subset E_x^{rk}$ will be called an interval in E_x^{rk} if $\{\mathcal{I}\}_\infty$ is an interval in $E_x^{(r)}$ for every $\alpha < \Omega_{r+k}$. Furthermore we shall say that a set E, $E \subset E_x^{rk}$, is perfect ($\mathcal{I}$), hypermeasurable ($\mathcal{I}$) if the set $\{E\}_\alpha$ is perfect ($\mathcal{I}$), hypermeasurable ($\mathcal{I}$) respectively.

If we now suppose that to every point

$$\bar{x} = \{\bar{x}_0, \bar{x}_1, \bar{x}_2, \cdots \bar{x}_\alpha, \cdots\}, \alpha < \Omega_{r+k},$$

of a set $E \subset E_x^{rk}$ there corresponds one and only one point

$$y = f(\bar{x}) = \{f_0(\bar{x}_0), f_1(\bar{x}_1), f_2(\bar{x}_2), \cdots f_\alpha(\bar{x}_\alpha), \cdots\}, \alpha < \Omega_{r+k},$$

of E_x^{rk}, then we have a single valued function $y = f(x)$ defined in $E \subset E_x^{rk}$.

Denote by $f(E)$ the set of all the points $y = f(x)$ where $x \epsilon E$. If $f(x_1) \neq f(x_2)$ in all cases when $x_1 \neq x_2$, $x_1 \epsilon E$, $x_2 \epsilon E$, then the function $f(x)$ is said to be *regular* in E.

A function $f(x)$ defined in E, $E \subset E_x^{rk}$, will be called continuous ($\mathcal{I}$), hypermeasurable ($\mathcal{I}$) in E if $\{f(x)\}_\alpha$ is respectively continuous ($\mathcal{I}$), hypermeasurable ($\mathcal{I}$) in $\{E\}_\alpha$ for every $\alpha < \Omega_{r+k}$. From this it follows that a function defined in $E \subset E_x^{rk}$ is hypermeasurable ($\mathcal{I}$) in E if E contains a perfect ($\mathcal{I}$) set $\mathcal{P}$ such that $f(x)$ is continuous ($\mathcal{I}$) in $\mathcal{P}$.

We denote by E_1 the set of all points

$$x = \{x_0, x_1, x_2, \cdots x_\alpha, \cdots\}, \alpha < \Omega_{r+k},$$

of E_x^{rk} having the following property: any one of the elements $x_0, x_1, x_2, \cdots x_\alpha, \cdots, \alpha < \Omega_{r+k}$, of $E_k^{(r)}$ is the rational point $\{2\}$ of $E_x^{(r)}$ and every other element is the rational point $\{1\}$ of $E_x^{(r)}$.

Let $|\bar{\alpha}|$ be the point

$$x = \{x_0, x_1, x_2, \cdots x_\alpha, \cdots\}$$

of E_x^{rk} such that $x_\alpha = \{2\}$ and $x_\alpha = \{1\}$ for $\alpha_1 \neq \alpha$.

We denote by E_2 the set of all points $\rho\, I$ where $\rho \epsilon E_x^{(r)}$ and $I = \{1, 1, 1, \cdots 1, \cdots\} \epsilon E_x^{rk}$.

Now we consider a correspondence

(7) $$\alpha = Z(\rho)$$

between the number α, $1 \leqq \alpha < \Omega_{r+k}$, and the point $\rho \in E_x^{(r)}$. It is easily seen that we can replace the correspondence (7) by the equivalent correspondence

(8) $$|\underline{\bar{\alpha}}| = Z(\rho I)$$

between the points

$$|\underline{\bar{\alpha}}|,\ \rho I$$

of E_x^{rk}.

We shall say that $f(x)$ is a function of Nicholas Parfentieff if

(i) $f(x)$ is regular in E_2;

(ii) $$E_1 \neq f(E_2).$$

We denote by (P) the class of all the functions of Nicholas Parfentieff.

THEOREM 1. *If a function $f(x)$ is continuous ($\mathfrak{I}$) and regular in $E_2 \subset E_x^{rk}$, then $f(x)$ is a function of N. Parfentieff.*

PROOF. Suppose the function $f(x)$ does not belong to the class (P). Since $f(x)$ is regular in E_2, we have: $E_1 = f(E_2)$. Since the function $y = f(x)$ is continuous ($\mathfrak{I}$) in E_2, to an interval $\mathfrak{I}(y_0) \subset E_x^{rk}$ $(y = f(x_0), x_0 \in E)$ there corresponds an interval $\mathfrak{I}(x_0)$ in such a manner that when x runs over the set $\mathfrak{I}(x_0)E_2$ then $f(x)$ runs over a set H contained in $\mathfrak{I}(y_0)\,f(E_2)$. We then write: $\mathfrak{I}(y_0) \rightarrow \mathfrak{I}(x_0)$. Let $\mathfrak{I}_\nu(y)$ be an interval in E_x^{rk} containing the point $|\underline{\bar{\nu}}| \in E_x^{rk}$. Evidently, to the interval $\mathfrak{I}_\nu(y)$ there corresponds an interval $\mathfrak{I}(x)$, $\mathfrak{I}_\nu(y) \rightarrow \mathfrak{I}(x)$, where $y = f(x)$. It is obvious that the set $\mathfrak{I}(x)E_2$ has the power $\geqq \aleph_0$, but the set $\mathfrak{I}_\nu(y)E_1$ consists of one and only one point $|\underline{\bar{\nu}}|$. Since the function $f(x)$ is regular in E_2, the set H of all the points $f(x)$ where $x \in \mathfrak{I}(x)E_2$ also has the power $\geqq \aleph_0$. However this is impossible in view of $H \subset \mathfrak{I}_\nu(y)E_1$.

THEOREM 2. *Every hypermeasurable ($\mathfrak{I}$) regular in E_2 function $y = f(x)$ is a function of N. Parfentieff.*

PROOF. Suppose the function $y = f(x)$ does not belong to the class (P). Since $f(x)$ is regular in E_2, then $E_1 = f(E_2)$. Since the function $f(x)$ is hypermeasurable ($\mathfrak{I}$) in E_2, there exists a perfect ($\mathfrak{I}$) set F contained in E_2 and such that $f(x)$ is continuous ($\mathfrak{I}$) in F. Let $E_1' = f(F)$ and let $|\underline{\bar{\mu}}|$ be any point of E_1'. It is clear that to an interval $\mathfrak{I}_\mu(y)$, $\mathfrak{I}_\mu(y) \subset E_x^{rk}$, $(y = f(x), x \in F)$ there corresponds an interval $\mathfrak{I}(x) \subset E_x^{rk}$ in such a manner that when x runs over the set $\mathfrak{I}(x)F$, then $y = f(x)$ runs over a set H' contained in $\mathfrak{I}_\mu(y)E_1'$; for $f(x)$ is regular in E_2 and since $\mathfrak{I}_\mu(y)E_1'$ consists of one and only one point. But the set $\mathfrak{I}(x)F$ has the power $\geqq \aleph_0$, and so H' has also the power $\geqq \aleph_0$. This is impossible since $H' \subset \mathfrak{I}_\mu(y)E_1'$.

Thus, in the mathematics $\mathfrak{L}$ of the hypermeasurable ($\mathfrak{I}$) functions we have: $2^{\aleph_r} \neq \aleph_{r+k}$ $(0 \leqq r < \Omega_0,\ 1 \leqq k < \Omega_0)$.

TCHEBOLSSARY, TCHAPAEWSKIPOSELOK, U.S.S.R.

Annals of Mathematics
Vol. 44, No. 1, January, 1943

LINEARLY ARC-WISE CONNECTED TOPOLOGICAL ABELIAN GROUPS

By R. C. James

(Received December 10, 1941)

Several types of connectedness of topological spaces can be defined in terms of continuous functions with real numbers as arguments and values in the space.[1] The function can be required to be linear if the space is a topological abelian group, giving rise to properties that can be used as necessary and sufficient conditions for the topological abelian group to be a linear topological space. If the concepts of convexity and boundedness are generalized to topological abelian groups, these properties lead to a theorem on normability analogous to one given by Kolmogoroff for linear topological spaces.[2] I am indebted to Dr. Michal for suggesting an investigation of linear arc-wise connectedness, and for his guidance.

1. Topological groups and linear topological spaces

By a topological group, we mean an abstract group which has a Hausdorff topology[3] with respect to which the group operations are continuous. A topological abelian group is a topological group whose abstract group is abelian. A linear topological space is a linear space[4] which has a Hausdorff topology with respect to which the fundamental operations $x + y$ and ax are continuous. It can easily be shown that a linear topological space is a topological abelian group and has the following properties:

I. If $ax = 0$, either $x = 0$ or $a = 0$.
II. If $ax = ay$ and $a \neq 0$, then $x = y$; if $ax = bx$ and $x \neq 0$, then $a = b$.
III. If U is an open set and $a \neq 0$, then aU is an open set.
IV. If U is a neighborhood of the origin, there is a neighborhood V of the origin such that $aV \subset U$ for all a satisfying $| a | \leqq 1$.

2. Continuous functions and arc-wise connectedness

The functions treated in this paper are exclusively functions with real numbers as arguments and values in a topological abelian group or a linear topological space. Such a function $f(t)$ is said to be continuous at $t = t_1$ if for any neighborhood U of $f(t_1)$ there is a $\delta > 0$ such that $f(t) \in U$ for all t satisfying $| t - t_1 | < \delta$. A continuous function will be called linear if it is additive [i.e. $f(t_1 + t_2) = f(t_1) + f(t_2)$] and continuous throughout its interval of definition. It will be convenient to introduce the concept of uniform continuity:

Definition 2.1. *A function $f(t)$ with real numbers as arguments and values in a topological group T is said to be uniformly continuous in an interval (a, b) if for any*

---

[1] See Pontrjagin (VI) for a discussion of these properties. (Roman numerals refer to the bibliography.)

[2] Kolmogoroff (IV).

[3] Hausdorff (II). The space satisfies (A), (B), (C), (5), pp. 228–229.

[4] Banach (1), pg. 26.

neighborhood U of the identity there is a positive number δ such that $f(t) \epsilon f(t_1) + U$ if t_1 and t belong to (a, b) and $| t_1 - t | < \delta$.

This definition can be readily extended to functions of several variables. It can be shown that a uniformly continuous function is continuous and that a function which is continuous in a closed interval is uniformly continuous in this interval.

Several types of arc-wise connectedness of topological spaces can be defined in terms of continuous functions of a real variable, two points x and y of the topological space being said to be joined by a curve or arc if there is a continuous function $f(t)$ such that $f(0) = x$ and $f(1) = y$.[5] Only one of these types of connectedness is needed in this paper (def. 2.2), although the others will be used for topological abelian groups in a more restricted sense as defined by means of linear functions.

DEFINITION 2.2. *A topological space is called simply connected if every closed path in the space can be continuously shrunk to a point; explicitly, if for every closed path $f(t)$ for which $f(0) = f(1)$ there is a function $g(s, t)$, continuous simultaneously in s and t, for which $g(0, t) \equiv f(t)$, $g(1, t) \equiv f(0)$, and $g(s, 0) \equiv g(s, 1) \equiv f(0)$.*

DEFINITION 2.3. *A topological abelian group T is called linearly arc-wise connected if: (1) the identity can be joined to any point of T by a curve defined by a linear function; or (2) any two points x_1 and x_2 of T can be joined by a curve defined by an affine function, i.e. a function $f(t)$ satisfying the functional relation: $f(0) + f(t_1 + t_2) = f(t_1) + f(t_2)$.[6] If there is only one linear path joining the identity to any given point of T, then T will be called uniquely linearly arc-wise connected.*

DEFINITION 2.4. *A topological abelian group T is called locally linearly arc-wise connected if: (1) for any neighborhood U of the identity there is a neighborhood V of the identity such that the identity can be joined to any point of V by a curve contained in U and defined by a linear function; or (2) for any neighborhood U' of a point x_1 of T there is a neighborhood V' of x_1 such that any point x_2 of V' can be joined to x_1 by a curve contained in U' and defined by an affine function.*

It can be shown that the two forms of definition 2.3 (or 2.4) are equivalent. From the fact that a connected topological group can be generated by an arbitrary neighborhood of the identity,[7] it follows that a topological abelian group which is connected and locally linearly arc-wise connected is also linearly arc-wise connected. A linear topological space can be shown to have the properties of definitions 2.2, 2.3, and 2.4. In fact, the linear path joining the identity to any point x is $f(t) \equiv tx$, and any closed path $f(t)$ can be continuously deformed to a point by means of the function $g(s, t) \equiv sf(0) + (1 - s)f(t)$. Property IV of section 1 shows that a linear topological space is locally linearly arc-wise connected.

[5] See Pontrjagin (VI), pp. 219–221, for explicit definitions.

[6] The definition of *curve* implies that $f(t)$ is also continuous.

[7] Pontrjagin (VI), pg. 76, theorem 15.

3. Points of finite order in a topological abelian group

Since a linear topological space has no non-zero points of finite order, this is a necessary condition for a topological abelian group to be a linear topological space. It will be shown in this section that a topological abelian group which is connected, simply connected and locally linearly arc-wise connected must have either no non-zero elements of finite order or the set of elements of the n^{th} order is dense in itself for each $n > 1$. However topological abelian groups can be found which are linearly arc-wise connected and locally linearly arc-wise connected and have any desired number of elements of a given prime order, provided an abstract abelian group exists which has this number of elements of that order. This is shown by the following examples:

Let T be composed of elements of the form $x = (x_1, x_2, \cdots)$, $y = (y_1, y_2, \cdots)$, where x_i and y_i are any complex numbers of absolute value 1. Define $x + y$ as $(x_1y_1, x_2y_2, \cdots)$ and a neighborhood U of x as the set of elements of the form $u = (u_1, u_2, \cdots)$, where $| u_i - x_i | < \epsilon$ for $i \leqq 1/\epsilon$. It can be shown that T is a topological abelian group which is linearly arc-wise connected and locally linearly arc-wise connected (but not simply connected) and the set of elements of the n^{th} order is dense in itself for each $n > 1$. It is an open question whether this is possible if the space is also simply connected. Let T' be composed of elements of the form $x = (x_1, x_2, \cdots, x_r)$, where x_i is any complex number of absolute value 1, and addition and neighborhoods are defined as above. If p is a prime, the element x is of the p^{th} order if and only if $x_j = \exp(2\pi i(n_j/p))$ for each j, and at least one n_j is not a multiple of p. Hence T' contains $p^r - 1$ elements of the p^{th} order, and any abstract abelian group contains $p^r - 1$ elements of the p^{th} order (for some integer r).

The following three lemmas will be used in establishing the theorem of this section:

LEMMA 3.1. *If a topological abelian group is linearly arc-wise connected and has a non-zero element of finite order, then for each positive integer n there are at least $n - 1$ non-zero elements x_i such that $nx_i = 0$.*

LEMMA 3.2. *If a topological abelian group is locally linearly arc-wise connected, and for some integer $p > 1$ there is an element of order p in every neighborhood of the identity, then this is true for any positive integer n, and the set of elements of order n is dense in itself (for each integer $n > 1$).*

Lemma 3.1 and the first part of lemma 3.2 can be readily proved by exhibiting certain closed linear paths which are not identically zero. Connectedness is not necessary in Lemma 3.2, since a closed linear path in some neighborhood of the identity is sufficient. The proof that the set of elements of order n ($n > 1$) is dense in itself if there is an element of order n in every neighborhood of the identity is as follows:

Let x be any element of order n and W a neighborhood of x. Then $W - x$ is an open set containing the identity. Let $U \subset W - x$ be a neighborhood of the identity which contains none of the elements ix ($i = 1, 2, \cdots, n - 1$), and

V be a neighborhood of the identity such that $V^n \subset U$.[8] There is by assumption an element y of order n belonging to V. Then $x + y \in W$ and $n(x + y) = nx + ny = 0$. If $i(x + y) = 0$ for $0 < i < n$, then $iy = -ix = (n - i)x$. But this is impossible, since $y \in V$, $iy \in V^i \subset U$, and $(n - i)x$ does not belong to U. Hence for each positive integer n the set of elements of order n is dense in itself.

LEMMA 3.3. *If a topological abelian group is locally linearly arc-wise connected and there is a positive integer n and a neighborhood U of the identity such that $x \in U$ and $nx = 0$ imply $x = 0$, then for any neighborhood M of the identity there is a neighborhood $N \subset M$ such that if $y \in x + N$, and x_1 is any point such that $nx_1 = x$, there is a unique point y_1 which belongs to $x_1 + N$ and is such that $ny_1 = y$.*

PROOF: For such a neighborhood U and any neighborhood M of 0, take a neighborhood V such that $V \subset M$ and $V \subset U$. Let N be the set of all points of V which can be joined to the identity by a linear path $f(t)$ for which there is a neighborhood L of the identity such that $f(t) + L \subset V$. Suppose $nx_1 = x$ and $y \in x + N$, or $y - x \in N$. Let $g(t)$ be the linear path joining the identity to $y - x$. Then $y_1 = x_1 + g(1/n) \in x_1 + N$, and $ny_1 = y$. Since $N \subset U$, there is no other point y_2 of $x_1 + N$ such that $ny_2 = y$.

It now remains to show that N is open. Let x be a point of N, $f(t)$ the linear path joining the identity to x, L the neighborhood of the identity such that $f(t) + L \subset V$, and L' a neighborhood of the identity such that $2L' \subset L$. Since T is locally linearly arc-wise connected, there is a neighborhood L'' of the identity such that if $y \in L''$ there is a linear path contained in L' which joins the identity to y. Then $L'' + x$ is an open set contained in N, for if $z \in L'' + x$, then $z - x \in L''$, and hence there is a linear path $g(t)$ contained in L' and joining the identity to x. Then $[g(t) + f(t)] + L' \subset V$ and $g(t) + f(t)$ joins the identity to z. Hence $z \in N$.

THEOREM 3.1. *If a topological abelian group T is connected, simply connected and locally linearly arc-wise connected, then T either has no non-zero elements of finite order, or else for each positive integer $n > 1$ the set of elements of the nth order is dense in itself.*

PROOF: Suppose there is a non-zero element of finite order and the points x_i for which $2x_i = 0$ is not dense in itself; that is, there is a neighborhood U of the identity such that $x \in U$ and $2x = 0$ imply $x = 0$. Since T is connected and locally linearly arc-wise connected, it is linearly arc-wise connected. Let $f(t)$ be a linear function of t for which $f(0) = 0$, $f(1) = x \neq 0$, where $x \in T$. By lemma 3.1, there is an element y of order two. Then $2[f(1/2) + y] = x$, and since T is linearly arc-wise connected, there is a function $F(t)$ which is continuous for $0 \leqq t \leqq 1$ and such that $F(0) = f(1/2)$, $F(1) = f(1/2) + y$. Also, $2F(t)$ is a closed path beginning and ending at x, and hence there is a function $g(s, t)$, continuous simultaneously in s and t, such that $g(0, t) \equiv 2F(t)$, $g(1, t) \equiv x$, $g(s, 0) \equiv g(s, 1) \equiv x$.

Take a neighborhood $N \subset U$ of the identity, as defined in lemma 3.3. There

[8] U^n is the set of all $y = u_1 + u_2 + \cdots + u_n$, where each u_i belongs to U.

is an ϵ_1 such that if $0 \leqq s < \epsilon_1$, $| t' - t | < \epsilon_1$, and $0 \leqq t \leqq 1$, then $g(s, t) \in N + g(0, t')$ for all t' $(0 \leqq t' \leqq 1)$.[9] Let $0 < s_1 < \epsilon_1$, and hence $g(s_1, t') \in g(0, t') + N$ for all t'. Since N has the properties of lemma 3.3, there is a unique point $x_1 \in N + F(t')$ such that $2x_1 = g(s_1, t')$. Define the function $F(s_1, t)$ by the relations $F(s_1, t') = x_1$. The function $F(s_1, t)$ is uniquely defined for all t $(0 \leqq t \leqq 1)$; and $F(s_1, 0) = F(0)$, $F(s_1, 1) = F(1)$, $2F(s_1, t) \equiv g(s_1, t)$.

It will now be shown that $F(s_1, t)$ is continuous in t. Let U be any neighborhood of $F(s_1, t')$. Since $F(s_1, t') + 0 \subset F(t') + N$, there is a neighborhood U_1 of $F(s_1, t')$ and a neighborhood V of 0 such that $U_1 + V \subset F(t') + N$ and $U_1 \subset U$. If δ_1 is such that $F(t') - F(t) \subset V$ for $| t' - t | < \delta_1$, then $U_1 + F(t') - F(t) \subset F(t') + N$, or $U_1 \subset F(t) + N$, for $| t' - t | < \delta_1$. Similarly, there is a δ_2 and a neighborhood U_2 of $g(s_1, t')$ such that $U_2 \subset g(0, t) + N$ if $| t' - t | < \delta_2$. Since $U_1 - F(s_1, t')$ and $U_2 - g(s_1, t')$ are both neighborhoods of the identity, there is a neighborhood M of the identity which is contained in their intersection. Take $N' \subset M$ and having the property of Lemma 3.3. Let δ be a positive number equal to or less than the smaller of δ_1 and δ_2, and such that if $| t' - t | < \delta$, then $g(s_1, t) \in g(s_1, t') + N'$. If $| t' - t | < \delta$, then there is a unique point x_1 in $F(s_1, t') + N'$ such that $2x_1 = g(s_1, t)$. But since $F(s_1, t)$ is the unique point in $F(t) + N$ such that $2F(s_1, t) = g(s_1, t)$, and $x_1 \in F(s_1, t') + N' \subset U_1 \subset F(t) + N$, it follows that $x_1 = F(s_1, t)$. That is, if $| t' - t | < \delta$, then $F(s_1, t) \in F(s_1, t') + N' \subset U_1 \subset U$. Hence $F(s_1, t)$ is a continuous function of t.

There is an ϵ (the same for all s_1, t_1) such that $g(s, t) \in g(s_1, t_1) + N$ if $| s_1 - s | < \epsilon$, $| t_1 - t | < \epsilon$, and s and t are in the interval $(0, 1)$. Hence the above process can be continued by "jumps" of ϵ. That is, if $F(s_1, t)$ is a continuous function of t and $F(s_1, 0) = F(0)$, $F(s_1, 1) = F(1)$, $2F(s_1, t) \equiv g(s_1, t)$; then a continuous function of t, $F(s, t)$, can be defined for any s $(s_1 < s < s_1 + \epsilon$, $0 \leqq s \leqq 1)$, such that $F(s, 0) = F(0)$, $F(s, 1) = F(1)$, and $2F(s, t) \equiv g(s, t)$. But $F(1, t)$ is a continuous function of t and $F(1, 0) = F(0)$, $F(1, 1) = F(1)$, and $2F(1, t) \equiv g(1, t) \equiv x$. This contradicts the fact that in each of the neighborhoods $F(0) + N$ and $F(1) + N$ of $F(0)$ and $F(1)$, respectively, there is only one point x_1 such that $2x_1 = x$. That is, the assumption that there is a point of finite order and a neighborhood U of the identity such that $x \in U$ and $2x = 0$ imply $x = 0$ has been shown to be impossible. Hence if there is a non-zero point of finite order, there is a point of the second order in every neighborhood of the identity. Then from Lemma 3.2 it follows that the set of points of order n is dense in itself (for each integer $n > 1$).

4. Conditions for a topological abelian group to be a linear topological space

The following lemma will be needed to establish the results of this section. It is equivalent to the continuity of multiplication by real numbers, as defined in the proof of theorem 4.1.

LEMMA 4.1. *If a topological abelian group T is uniquely linearly arc-wise*

---

[9] See definition 2.1 and the following discussion.

connected and locally linearly arc-wise connected, then the function $F(t, x) = f(t)$, *where* $f(t)$ *defines the linear path joining the identity to* x, *is simultaneously continuous in* t *and* x.

PROOF: Let x' and t' be any values of x and t, and W a neighborhood of $f(t') = F(t', x')$, where $f(t)$ is the linear path joining the identity to x. Since $0 + f(t') = f(t')$, there are neighborhoods V_1 and V_2 of the identity and $f(t')$, respectively, such that $V_1 + V_2 \subset W$. Since T is locally linearly arc-wise connected, there is a neighborhood V' of the identity such that each point of V' can be joined to the identity by a linear path contained in V_1. Take $U = x' + V'$ and $x \in U$. Let $g(t)$ and $f(t)$ be the linear paths joining the identity to x and x', respectively. Then $g(t) - f(t)$ is the linear path joining the identity to $x - x' \in V'$, and is therefore contained in V_1. Since $f(t)$ is continuous, there is an $\epsilon > 0$ such that $f(t) \in V_2$ for all t such that $| t' - t | < \epsilon$. If $| t' - t | < \epsilon$, then $g(t) - f(t) + f(t) = g(t)$ belongs to W. Since $g(t)$ was the linear path joining the identity to an arbitrary point x of U, it then follows that $F(t, x)$ is simultaneously continuous in t and x.

THEOREM 4.1. *A necessary and sufficient condition that a topological abelian group* T *be a linear topological space is that* T *be uniquely linearly arc-wise connected and locally linearly arc-wise connected.*

PROOF: For each real number a and point x of T, define $a \cdot x$ as $F(a, x) = f(a)$, where $f(t)$ is the linear function joining the identity to x.[10] Lemma 4.1 states that this multiplication is continuous simultaneously in a and x. From the fact that $f(t)$ is additive, it also follows that the multiplication satisfies the postulates for a linear space. Hence T is a linear topological space.

Several variations of this theorem can be gotten by giving other conditions in place of the unique linear arc-wise connectedness. It can be shown that if the linear path joining the identity to some point x is unique, or if there are no non-zero points of finite order, then the topological abelian group is uniquely linearly arc-wise connected if it is linearly arc-wise connected.

COROLLARY. *A necessary and sufficient condition that a topological abelian group* T *be a linear topological space is that* T *be connected, locally linearly arc-wise connected, and for some element* $x \in T$ *there is only one linear path joining the identity to* x.

COROLLARY. *A necessary and sufficient condition that a topological abelian group be a linear topological space is that it be connected, locally linearly arc-wise connected, and possess no non-zero elements of finite order.*

If a connected topological abelian group T is locally linearly arc-wise connected, it is also linearly arc-wise connected. If T is also simply connected, it then follows from theorem 3.1 that T either contains no non-zero elements of finite order, or the set of elements of order n is dense in itself for each $n > 1$. Hence the second corollary above can be written in the form:

THEOREM 4.2. *A necessary and sufficient condition that a topological abelian*

---

[10] $f(t)$ is only required to be defined for $0 \leq t \leq 1$, but it can be uniquely extended to all real numbers by using the additive property.

group be a linear topological space is that it be connected, simply connected, locally linearly arc-wise connected, and there exist an integer $n > 1$ and a neighborhood U of the identity such that U contains no elements of order n.

5. Convex topological groups

A. Kolmogoroff and A. Tychonoff have called a neighborhood U of the origin of a linear topological space convex if $ax + (1 - a)y$ is in U for any elements x and y of U and real number a in the interval $(0, 1)$.[11] A linear topological space is said to be *locally convex* if for any neighborhood U of the origin there is a convex neighborhood V contained in U. John von Neumann calls a linear topological space *convex* if for any neighborhood U of the origin there is a neighborhood $V \subset U$ such that $2V = V^2$.[12] J. V. Wehausen has shown that these concepts of convexity and local convexity are equivalent.[13] Convexity can be extended to topological groups in a number of ways, the following seeming to be the most satisfactory.

DEFINITION 5.1. *A topological group is called convex if for every neighborhood U of the identity there is a neighborhood $V \subset U$ such that $nx \in V^n$ implies $x \in V$ (for every positive integer n).*

When the topological group is a linear topological space, this definition is equivalent to the above definitions of convexity and local convexity. It is also evident that a convex topological group has no non-zero points of finite order. The definition of convexity given by von Neumann can be applied to topological groups without rewording, but is not entirely satisfactory, since for each y in some neighborhood of the identity it implies the existence of an x such that $2x = y$. Convexity as given in definition 5.1 also has the advantage of being a consequence of normability (def. 6.4).

LEMMA 5.1. *A topological abelian group T which is convex and linearly arc-wise connected is locally linearly arc-wise connected.*

PROOF: Take any neighborhood U of the identity. Choose a neighborhood V such that $\bar{V} \subset U$ and $nx \in V^n$ implies $x \in V$. Let x be any point of V and $f(t)$ the linear path joining the identity to x: $f(0) = 0, f(1) = x$. Let p/q be any rational number in the interval $(0, 1)$, where p and q are positive integers. Then $qf(p/q) = f(p) \in V^p \subset V^q$. Hence $f(p/q) \in V$. Since $f(t)$ is continuous, it then follows that $f(t) \in \bar{V}$ for $0 \leqq t \leqq 1$. That is, each point of V can be joined to the identity by a linear arc contained in U, and hence T is locally linearly arc-wise connected.

THEOREM 5.1. *A necessary and sufficient condition that a topological abelian group be a convex linear topological space is that it be convex and linearly arc-wise connected.*

This theorem is an immediate consequence of the above lemma and theorem 4.1.

---

[11] Kolmogoroff (IV) and Tychonoff (VII).

[12] Neumann (V), pg. 4. $2V$ is the set of all $u = v + v$, where $v \in V$; $V^2 = V + V$ is the set of all $v_1 + v_2$, where v_1 and v_2 belong to V.

[13] Wehausen (VIII), pg. 158.

6. Bounded sets and normable topological abelian groups

Several equivalent definitions of bounded sets in linear topological spaces have been given. The following one is due to J. v. Neumann, and can be extended to topological groups. This property will be used in finding conditions for normability of topological abelian groups.

DEFINITION 6.1. *A set S of a linear topological space is called bounded if for any neighborhood U of the origin there is a number a such that $S \subset aU$.*[14]

DEFINITION 6.2. *A set S of a topological group will be called bounded if for any neighborhood U of the identity there exists an integer n such that $(1/m)S \subset U$ for all integers $m \geqq n$, where $(1/m)S$ is the set of all x such that $mx \in S$.*

DEFINITION 6.3. *A topological group will be called locally bounded if it has a bounded neighborhood of the identity.*

It can easily be shown that definition 6.2 is equivalent to definition 6.1 when the topological group is a linear topological space. Definition 6.1 could have been revised for topological groups by merely requiring that a be an integer. However this has the disadvantage of implying that for each element x of a bounded set $S \subset aU$ (where U is a neighborhood of the identity) of a topological group there exists an element u of U such that $x = au$. For example, if this definition were used, no non-zero elements of the multiplicative group of rational numbers would be bounded. Boundedness as given in definition 6.2 also has the advantage that local boundedness is implied by normability (def. 6.4), while the above example shows that this would not be true for the other form.

Even if the topological group were linearly arc-wise connected, def. 6.2 would not be equivalent to the form gotten by replacing "$(1/m)S \subset U$" by "$S \subset mU$". This is illustrated by the multiplicative group of all complex numbers with absolute value 1, for any single element is bounded in the revised sense, but no element is bounded in the sense of def. 6.2. This difference is closely related to the fact that the form "$(1/m)S \subset U$" implies that the set of all nx (for n an integer) is unbounded (symbolically, $\lim_{n\to\infty} nx = \infty$) if $x \neq 0$, while the second does not. Requiring that the relation $(1/m)S \subset U$ hold for all $m \geqq n$, rather than merely $(1/n)S \subset U$, has easily seen advantages, but becomes ambiguous when the topological group is a linear topological space.

DEFINITION 6.4. *A topological abelian group T is normable if to each point x of T there can be associated a non-negative real number $| x |$ which satisfies the conditions:*

(1). $| nx | = | n || x |$ *for all integers n and elements x of T.*

(2). $| x + y | \leqq | x | + | y |$.

(3). *The system of neighborhoods of T is topologically equivalent to the system of neighborhoods defined in terms of the distance between two points, as given by $\rho(x, y) = | x - y |$.*

THEOREM 6.1. *A necessary and sufficient condition that a linearly arc-wise*

[14] This and three other equivalent definitions are given in Hyers (III).

connected topological abelian group T be normable is that T be a normable linear topological space.[15]

PROOF: It is evident that the conditions of the theorem are sufficient. Now suppose that T is a normed linearly arc-wise connected topological abelian group. It follows from (1) and (3) of definition 6.4 that $|x| = 0$ if and only if $x = 0$. T then contains no non-zero points of finite order, for if $nx = 0$ it follows that $|nx| = 0 = |n||x|$, and hence that $x = 0$. It will now be shown that T is also locally linearly arc-wise connected. Let U be the neighborhood of the identity consisting of all points x for which $|x| < \epsilon_1$, and V be the neighborhood of the identity consisting of all points x for which $|x| < \epsilon_2$, where $\epsilon_2 < \epsilon_1$. Let y be a point of V and $f(t)$ the linear path joining the identity to y $[f(0) = 0, f(1) = y]$. Suppose $f(t)$ does not belong to V for some rational number $t_1 = p/q$, $0 < p/q < 1$, where p and q are positive integers. Then $|f(p/q)| \geqq \epsilon_2$, and $|f(p)| = q|f(p/q)| \geqq q\epsilon_2$. But since $|f(1)| < \epsilon_2$, it follows that $|f(p)| < p\epsilon_2$. These statements are contradictory, since $0 < p/q < 1$, or $q > p$. Hence $f(t) \epsilon V$ for all rational values of t in the interval $0 \leqq t \leqq 1$. From the continuity of $f(t)$ it follows that $f(t) \epsilon \bar{V} \subset U$ for all t. It has thus been shown that T is locally linearly arc-wise connected. T is then a linear topological space by the second corollary of theorem 4.1. It only remains to show that $|ax| = |a||x|$ for all real a. From definition 6.4, $|px| = |p||x|$ for all integers p. Hence $|(p/q)x| = |p||x/q| = |p/q||q||x/q| = |p/q||x|$. That is, $|ax| = |a||x|$ for all rational a. From the continuity of ax and the continuity of the norm, it now follows that $|ax| = |a||x|$ for all real a.

THEOREM 6.2. *A necessary and sufficient condition for the normability of a linearly arc-wise connected topological abelian group T is that T be convex and locally bounded.*

PROOF. If T is a normable, linearly arc-wise connected topological abelian group, then T is a normable linear topological space. If the system of neighborhoods is defined in terms of the norm, it is easily seen that each neighborhood is bounded and convex. Conversely, if the topological abelian group T is convex and linearly arc-wise connected, it follows from the second corollary of theorem 4.1 that T is a linear topological space. If T is also locally bounded, it follows from a theorem of Kolmogoroff's[16] that T is a normable linear topological space and hence a normable topological abelian group.

CALIFORNIA INSTITUTE OF TECHNOLOGY.

BIBLIOGRAPHY

I. S. BANACH, *Théorie des Opérations Linéaires*, Warsaw (1932).
II. F. HAUSDORFF, *Mengenlehre*, Berlin (1927).

[15] A linear topological space is normable under the same conditions as given in definition 6.4, except that (1) is replaced by $|ax| = |a||x|$ for all real numbers a and points x of T. See Kolmogoroff (IV), pg. 30.

[16] Kolmogoroff (IV). "*A necessary and sufficient condition for the normability of a linear topological space is the existence of a bounded convex open set.*"

III. D. H. Hyers, *Locally bounded linear topological spaces*, Revista de Ciencias, vol. 41, pp. 555–574.

IV. A. Kolmogoroff, *Zur Normierbarkeit eines allgemeinen topologischen linearen Raumes*, Studia Mathematica, vol. 5, pp. 29–33.

V. John von Neumann, *On Complete Topological Spaces*, Transactions of the American Mathematical Society, vol. 37, pp. 1–20.

VI. I. L. Pontrjagin, *Topological Groups*, Princeton (1939).

VII. A. Tychonoff, *Ein Fixpunktsatz*, Math. Annalen, vol. 111, pp. 767–776.

VIII. John V. Wehausen, *Transformations in linear topological spaces*, Duke Mathematical Journal, vol. 4, pp. 157–169.

Annals of Mathematics
Vol. 44, No. 1, January, 1943

CONTRIBUTIONS TO THE ANALYTIC THEORY OF CONTINUED FRACTIONS AND INFINITE MATRICES

By E. D. Hellinger and H. S. Wall

(Received June 16, 1942)

1. Introduction

In this paper we study continued fractions of the form

$$\frac{1}{b_1 + z} - \frac{a_1^2}{b_2 + z} - \frac{a_2^2}{b_3 + z} - \cdots, \tag{1.1}$$

in which the a_p are *real and positive*, and the b_p are *complex numbers with nonnegative imaginary parts*. In addition, we are concerned with certain infinite matrices connected with this continued fraction.

In the earlier investigations of continued fractions of this form, beginning with the classical work of Stieltjes [15][1], the b_p have been supposed *real*. For this case, and with some additional restrictions, Stieltjes was able to connect the continued fraction with one or more integrals of the form

$$\int_0^\infty \frac{d\phi(u)}{z + u}. \tag{1.2}$$

For particular cases where these restrictions are relaxed, Van Vleck [16] obtained again a connection with an analogous integral the range of which he had to extend over the whole real axis. Hilbert's [11] famous theory of bounded quadratic forms in which the ideas of Stieltjes are in the background allows immediate application to continued fractions of the form (1.1) and their connection with integrals of the form (1.2) but with a finite range of integration.[2] Grommer [4] showed that the process of Hilbert can be applied to more general cases where the integral extends from $-\infty$ to $+\infty$. A general theory of the continued fraction (1.1) in which the a_p and b_p are real and $a_p \neq 0$ was first developed by Hamburger [5] following the pattern laid down by Stieltjes. At about the same time the general case was treated by several other mathematicians: Hellinger [6] employed Hilbert's theory of infinite linear systems, R. Nevanlinna [12] used methods of function theory and asymptotic series, and Carleman [2] used his theory of integral equations.

In contrast with earlier extensions of Stieltjes' original theory, the problem of the present paper is unsymmetrical from two points of view: the functions which are obtained exist in general only in the upper half-plane, $\mathfrak{I}(z) > 0$, and the infinite matrix J of the system of equations

$$-a_p x_p + (b_{p+1} + z)x_{p+1} - a_{p+1}x_{p+2} = 0,\ p = 0, 1, 2, \cdots, (a_0 = 0), \tag{1.3}$$

Presented to the American Mathematical Society, April, 18, 1942.

[1] Numbers in square brackets refer to the bibliography.

[2] Cf. Hellinger-Toeplitz [8].

is not Hermitian. In spite of this asymmetry, we are able to develop a theory analogous to the classical theory. The principal points may be summarized as follows.

A. The nest of circles.[3] Regarding the continued fraction as an infinite product of linear transformations, we find for $\mathfrak{I}(z) > 0$ a nest of circles $K_1(z)$, $K_2(z)$, $K_3(z)$, $\cdots$ each inside the preceding and lying in the lower half-plane, such that the n-th approximant of (1.1) is on $K_n(z)$. We find for the radius $r_n(z)$ of $K_n(z)$ the formula

$$r_n(z) = \frac{1}{2 \sum_{p=1}^{n} \mathfrak{I}(b_p + z) \mid B_{p-1}(z) \mid^2},$$

(cf. (2.16) and (2.13)) where $x_p = B_{p-1}(z)$ is that solution of (1.3) for which $x_1 = 1$. We have to distinguish two cases according as $r(z) = \lim_{n=\infty} r_n(z)$ is positive (*limit-circle case*) or zero (*limit-point case*). In the latter case, (1.1) converges, while in the former case it may converge or diverge (§2).

B. Theorem of invariability. We show that the distinction between the two cases is independent of the particular value of z. This is accomplished in two different ways. 1. Using methods of function theory, particularly the *Stieltjes-Vitali* theorem, we show that it is sufficient to consider the continued fraction (1.1) for $z = 0$. *In this way we find that in addition to the distinction between the two cases, also the fact of convergence or divergence of* (1.1) *is independent of the particular value of* z (§ 4, 5). 2. Using for the linear difference equations an idea which Weyl [19][4] has applied in similar problems on differential and difference equations, we prove a more general formulation of the theorem of invariability (Theorem 7.1).

C. Reciprocals of the matrix J.[5] The matrix J always has at least one bounded reciprocal. However, while for real b_p the limit-point case is characterized by the fact that there is just one bounded reciprocal, there may be infinitely many for complex b_p. Therefore, we introduce a more restrictive boundedness condition (E-boundedness)[6] and obtain an analogous characterization, for the limit-point case (§6).

D. Asymptotic representation. The nest of circles yields a necessary and sufficient condition for an arbitrary function $f(z)$ to be asymptotically equal (Definition 8.1) to the continued fraction (§8).

---

[3] These circles have been applied for the case of real b_p by Hellinger [6]. Earlier applications in related problems were made by Weyl [18] and Hamburger [5]. Cf. also R. Nevanlinna [12], H. Weyl [19], and Paydon-Wall [13].

[4] The use of this idea which extends our original theorem and simplifies the proof was kindly suggested to us by the referee. Cf. further footnote 15.

[5] This theory has been given for the case of real a_p, b_p by Hellinger [6]; cf. also Beth [1].

[6] This amounts only to introducing weight factors in the sum of squares of the variables in the usual definition of boundedness. This has often been done ever since the start of the theory (cf. e.g. [9], §12). For problems somewhat related to our problem one will find an analogous change of the condition of convergence (boundedness) in Weyl [19], in the space of infinitely many variables as well as in the space of functions.

E. Stieltjes integral representation. If for $\Im(z) > 0$ the values of an analytic function $f(z)$ lie in the circle $K_n(z)$ for every n, then $f(z)$ has an integral representation of the form (1.2) with range of integration in general extended over the entire real axis, $\phi(u)$ being still a real nondecreasing function. Unlike the case where the b_p are real, the moments

$$\int_{-\infty}^{+\infty} u^p d\phi(u), \qquad p = 0, 1, 2, \cdots,$$

do not necessarily exist in the general case (§8).

PART I

CONTINUED FRACTIONS AND LINEAR TRANSFORMATIONS IN ONE VARIABLE

2. The nest of circles

A continued fraction

$$\frac{c_1}{d_1} - \frac{c_2}{d_2} - \frac{c_3}{d_3} - \cdots \tag{2.1}$$

may be regarded as an "infinite product" of the linear fractional transformations

$$t_n(w) = c_n/(d_n - w), \qquad n = 1, 2, 3, \cdots.$$

The case where these transformations carry some circular region H into all or a part of itself

$$t_n(H) \subset H \tag{2.2}$$

is of particular interest. For if in this case we put $t_1t_2 \cdots t_n(H) = H_n$, $n = 1, 2, 3, \cdots$, then $H \supset H_1 \supset H_2 \supset H_3 \supset \cdots$, and we are therefore in possession of important facts concerning the values of the "generalized approximants" $t_1t_2 \cdots t_n(w)$ of (2.1) when the parameter w lies in H.

We shall be concerned with the sequence of transformations

$$t_n(w) = \frac{a_{n-1}a_n^{-1}}{a_n^{-1}b_n - w}, \qquad n = 1, 2, 3, \cdots, \tag{2.3}$$

where $a_0 = 1, a_1, a_2, \cdots$ are *positive real numbers*, and $b_1, b_2, b_3, \cdots$ are complex numbers

$$b_n = \Re(b_n) + i\Im(b_n) \tag{2.4}$$

with nonnegative imaginary part

$$\Im(b_n) = \beta_n \geqq 0, \qquad n = 1, 2, 3, \cdots. \tag{2.5}$$

If H is the lower half-plane, $\Im(w) \leqq 0$, then under these conditions the transformation $t = t_n(w)$ from (2.3) satisfies the condition (2.2) because $a_{n-1}a_n^{-1}$ is real and positive and $\Im(a_n^{-1}b_n) \geqq 0$.

The continued fraction generated by these transformations is

$$\frac{a_1^{-1}}{a_1^{-1}b_1} - \frac{a_1 a_2^{-1}}{a_2^{-1}b_2} - \frac{a_2 a_3^{-1}}{a_3^{-1}b_3} - \cdots. \tag{2.6}$$

The notation has been chosen in such a way as to simplify and lend symmetry to the formulas which will be employed. The continued fraction is of course equivalent to

$$\frac{1}{b_1} - \frac{a_1^2}{b_2} - \frac{a_2^2}{b_3} - \cdots.$$

We may write $t_1t_2 \cdots t_n(w)$ in the form

$$t = t_1 t_2 \cdots t_n(w) = \frac{A_n - A_{n-1}w}{B_n - B_{n-1}w}, \qquad n = 1, 2, 3, \cdots. \tag{2.7}$$

The A_n, B_n are independent of w and satisfy the recurrence formulas

$$\begin{aligned} a_{n+1}A_{n+1} &= b_{n+1}A_n - a_nA_{n-1}, \\ a_{n+1}B_{n+1} &= b_{n+1}B_n - a_nB_{n-1}, \end{aligned} \qquad n = 0, 1, 2, \cdots, \tag{2.8}$$

where

$$A_{-1} = -1, \quad B_{-1} = 0, \quad A_0 = 0, \quad B_0 = 1, \quad A_1 = a_1^{-1}, \quad B_1 = a_1^{-1}b_1. \tag{2.9}$$

The determinant of the transformation (2.7) may be readily shown to be given by

$$A_nB_{n-1} - A_{n-1}B_n = a_n^{-1}, \qquad n = 1, 2, 3, \cdots. \tag{2.10}$$

The inverse transformation is

$$w = t_n^{-1}t_{n-1}^{-1} \cdots t_1^{-1}(t) = \frac{A_n - B_n t}{A_{n-1} - B_{n-1}t}. \tag{2.11}$$

The quotient A_n/B_n is the n-th approximant of the continued fraction (2.6). By (2.7) we have: $A_n/B_n = t_1t_2 \cdots t_n(0) = t_1t_2 \cdots t_{n+1}(\infty)$.

From the second recursion formula (2.8) we get

$$a_nB_n\bar{B}_{n-1} = b_n \mid B_{n-1} \mid^2 - a_{n-1}B_{n-2}\bar{B}_{n-1}, \qquad n = 1, 2, 3, \cdots,$$

or if we put

$$q_n = \mathfrak{J}(B_n\bar{B}_{n-1}), \qquad n = 1, 2, 3, \cdots, \tag{2.12}$$

and consider only the imaginary part, we find

$$a_nq_n = \beta_n \mid B_{n-1} \mid^2 + a_{n-1}q_{n-1}, \qquad n = 1, 2, 3, \cdots,$$

and therefore, using (2.9),

$$a_n q_n = \sum_{p=1}^{n} \beta_p \mid B_{p-1} \mid^2, \qquad n = 1, 2, 3, \cdots. \tag{2.13}$$

If we recall that $B_0 = 1$ and that $\beta_p \geqq 0$ and if we assume in addition that

$$\Im(b_1) = \beta_1 > 0, \tag{2.14}$$

we can conclude that $q_n > 0$. Therefore, (2.12) shows that $B_n \neq 0$.

Next we determine the region H_n into which the lower half-plane $\Im(w) \leqq 0$ is carried by one of the transformations (2.7). Since $B_n \neq 0$, $B_{n-1} \neq 0$, this transformation maps the real w-axis into a proper circle K_n of the t-plane which bounds H_n. Inasmuch as it maps the point $w = B_n/B_{n-1}$ into $t = \infty$ and the center $t = C_n$ of K_n can be produced by inversion of $t = \infty$ in K_n, then C_n must correspond under this transformation to the reflection $\bar{w} = \bar{B}_n/\bar{B}_{n-1}$ of $w = B_n/B_{n-1}$ in the real w-axis:

$$C_n = \frac{A_n \bar{B}_{n-1} - A_{n-1}\bar{B}_n}{B_n \bar{B}_{n-1} - B_{n-1}\bar{B}_n}. \tag{2.15}$$

Since $t_1 t_2 \cdots t_n(0) = A_n/B_n$ lies upon K_n we then find that the radius of K_n is

$$r_n = \left| \frac{A_n}{B_n} - C_n \right| = \frac{1}{2a_n q_n}. \tag{2.16}$$

In particular, the circle K_1 has the center $C_1 = -i/2\beta_1$ and the radius $r_1 = 1/2q_1 = 1/2\beta_1$, and therefore is tangent to the real axis from below at the origin. Since we have proved $H_1 \supset H_2 \supset H_3 \supset \cdots$ we conclude that for all points of these domains

$$-\frac{1}{\beta_1} \leqq \Im(w) \leqq 0, \qquad |w| \leqq \frac{1}{\beta_1}. \tag{2.17}$$

We have to distinguish two cases:

Case I. The limit-point case. The circular regions H_n have one and only one point f in common; the radius r_n of the circle K_n has the limit 0 for $n = \infty$; the infinite series

$$\sum_{p=1}^{\infty} \beta_p \, |B_{p-1}|^2 \tag{2.18}$$

diverges.

Case II. The limit-circle case. The circular regions H_n have a circular region in common; the circle bounding this common circular region has radius

$$r = \lim_{n=\infty} r_n > 0;$$

the infinite series (2.18) converges.

In Case I we have, *uniformly* for all w in H:

$$\lim_{n=\infty} t_1 t_2 \cdots t_n(w) = f.$$

Inasmuch as $w = 0$ is in H it follows that $\lim_{n=\infty}(A_n/B_n) = f$. Therefore, since the denominators B_n are different from 0, the continued fraction (2.6) is convergent in this case, and its value is f. The continued fraction will be called

completely convergent[7] in Case I. In Case II the continued fraction may converge or diverge; if it converges we shall say that it is *simply convergent.*

3. Complete convergence

We shall give two theorems which furnish sufficient conditions for the complete convergence of the continued fraction (2.6).

THEOREM 3.1. *If* $\liminf a_n$ *is finite, and* $\mathfrak{J}(b_n) \geqq k > 0$, $n = 1, 2, 3, \cdots$, *where* k *is a positive constant, then the continued fraction* (2.6) *is completely convergent.*[8]

PROOF. We prove the equivalent statement: If the series (2.18) converges and if

$$\mathfrak{J}(b_p) = \beta_p \geqq k > 0, \tag{3.1}$$

then $\lim a_n = \infty$. Indeed, from the convergence of (2.18) it follows that $\lim_{p=\infty}\beta_p \mid B_{p-1} \mid^2 = 0$, so that, on account of (3.1), $\lim_{p=\infty} B_{p-1} = 0$, and

$$\lim_{p=\infty} \mathfrak{J}(B_p \bar{B}_{p-1}) = 0. \tag{3.2}$$

On the other hand, we conclude from (2.13) and (3.1) that

$$a_n \mathfrak{J}(B_n \bar{B}_{n-1}) = \sum_{p=1}^{n} \beta_p \mid B_{p-1} \mid^2 \geqq k \mid B_0 \mid^2 = k > 0,$$

which, with (3.2), gives $\lim_{n=\infty} a_n = \infty$.

The second theorem concerns the continued fraction

$$\frac{1}{b_1} - \frac{1}{b_2} - \frac{1}{b_3} - \cdots \tag{3.3}$$

which is obtained by letting $a_n = 1$, $n = 1, 2, 3, \cdots$, in (2.6). As a matter of fact, this is no essential specialization. We note first of all that (3.3) necessarily diverges if all the b_n with odd subscripts vanish. For in this case $B_1 = B_3 = B_5 = \cdots = 0$. Suppose that $b_1 = b_3 = b_5 = \cdots = b_{2n-1} = 0$, $b_{2n+1} \neq 0$. Then, inasmuch as

$$t_1 t_2 \cdots t_{2n+1}(w) = -b_2 - b_4 - \cdots - b_{2n} + \frac{1}{b_{2n+1} - w},$$

it is clear that (3.3) is equivalent to a continued fraction of the same form but with the term $(-b_2 - b_4 - \cdots - b_{2n})$ added on, and with the first partial denominator different from 0.

THEOREM 3.2. *Suppose that* $b_1 \neq 0$, $\mid \mathfrak{R}(b_p) \mid \leqq k\mathfrak{J}(b_p)$, $p = 1, 2, 3, \cdots$, *where* k *is a positive constant, so that* $\mathfrak{J}(b_1) = \beta_1 > 0$. *Then the continued fraction*

[7] For real w this notion has been introduced by Hamburger [5].

[8] See Hellinger [6] for the case where $\mathfrak{J}(b_n)$ is independent of n.

(3.3) *is completely convergent if the series* $\sum |b_p|$ *diverges, and is divergent if this series converges.*[9]

PROOF. If $\sum |b_p|$ diverges, then the inequality $|b_p| \leqq (1+k)\beta_p$, $\beta_p = \mathfrak{J}(b_p)$, shows that $\sum \beta_p$ diverges. Since $B_p \neq 0$, $p = 0, 1, 2, \cdots$, we conclude from (2.8) that

$$\left|\frac{B_p}{B_{p-2}}\right| = \left|\frac{b_p B_{p-1} - B_{p-2}}{B_{p-2}}\right| \leqq 1 + \left|\frac{b_p B_{p-1}}{B_{p-2}}\right|$$

$$= 1 + \frac{|b_p|\cdot|B_{p-1}|^2}{|B_{p-1}\bar{B}_{p-2}|} \leqq 1 + \frac{(1+k)\beta_p|B_{p-1}|^2}{q_{p-1}}.$$

Therefore, since by (2.13) $\{q_n\}$ is a monotone nondecreasing sequence and $q_1 = \beta_1 > 0$:

$$\left|\frac{B_p}{B_{p-2}}\right| \leqq 1 + \frac{(1+k)}{\beta_1}(q_p - q_{p-1}) \leqq 1 + \frac{(1+k)}{\beta_1}(q_p - q_{p-2}),$$
$$p = 2, 3, 4, \cdots,$$

so that

$$|B_p| \leqq |B_{p-2}|\, e^{c_1}(q_p - q_{p-2}) \leqq c_2 e^{c_1 q_p},$$

where c_1, c_2 are positive constants. Now $|B_p B_{p-1}| \geqq q_p > \beta_1$, so that

$$|B_{p-1}| \geqq \frac{\beta_1}{|B_p|} \geqq c_3 e^{-c_1 q_p},$$

c_3 being a positive constant, or $\beta_p |B_{p-1}|^2 e^{2c_1 q_p} \geqq c_3^2 \beta_p$. Summing over p from 1 to n we get

$$\sum_{p=1}^{n} \beta_p |B_{p-1}|^2 e^{2c_1 q_p} \geqq c_3^2 \sum_{p=1}^{n} \beta_p;$$

or, since $q_p \geqq q_{p-1}$:

$$q_n e^{2c_1 q_n} = e^{2c_1 q_n} \sum_{p=1}^{n} \beta_p |B_{p-1}|^2 \geqq c_3^2 \sum_{p=1}^{n} \beta_p.$$

Inasmuch as $\sum \beta_p$ diverges, we conclude that $\lim_{n\to\infty} q_n = \infty$, and therefore Case I holds.

If $\sum |b_p|$ converges the continued fraction diverges by a well-known theorem (cf., e.g. Perron [14] p. 235).

4. Theorem of invariability

We consider now the continued fraction obtained from (2.6) by replacing b_p by $b_p + z$, $p = 1, 2, 3, \cdots$, namely:

$$(4.1) \qquad \frac{a_1^{-1}}{a_1^{-1}(b_1+z)} - \frac{a_1 a_2^{-1}}{a_2^{-1}(b_2+z)} - \frac{a_2 a_3^{-1}}{a_3^{-1}(b_3+z)} - \cdots.$$

---

[9] This theorem, except for the notion of complete convergence, has been given first by Van Vleck [17], p. 229, Theorem 6.

We keep the conditions

$$a_n > 0,\ \mathfrak{I}(b_n) = \beta_n \geqq 0, \qquad n = 1, 2, 3, \cdots, \tag{4.2}$$

and drop the condition $\mathfrak{I}(b_1) = \beta_1 > 0$. If we suppose that

$$\mathfrak{I}(z) = y > 0 \tag{4.3}$$

then the preceding theory will hold for (4.1) because $\mathfrak{I}(b_1 + z) = \beta_1 + y \geqq y > 0$. We shall use the notation of §2 except that dependence upon z will be indicated in the customary manner. In particular, the denominators will be denoted by $B_n(z)$; they satisfy the recursion formulas (2.8) which now read

$$-a_nB_{n-1}(z) + (b_{n+1} + z)B_n(z) - a_{n+1}B_{n+1}(z) = 0, \qquad n = 0, 1, 2, \cdots. \tag{4.4}$$

This is a system of infinitely many linear equations in $B_0(z)$, $B_1(z)$, $B_2(z)$, $\cdots$ with the following matrix of coefficients:

$$J = \begin{pmatrix} b_1 + z, & -a_1, & 0, & 0, & \cdots \\ -a_1, & b_2 + z, & -a_2, & 0, & \cdots \\ 0, & -a_2, & b_3 + z, & -a_3, & \cdots \\ 0, & 0, & -a_3, & b_4 + z, & \cdots \\ & \cdots & \cdots & \cdots & \end{pmatrix}. \tag{4.5}$$

Matrices of this form in which the coefficients $a_n \neq 0$ have been considered in the theory of infinite matrices under the name of *J-matrices* [8]. Since the theory of these matrices is equivalent to the theory of the continued fractions (4.1) it is appropriate to call these continued fractions *J-fractions*.

The J-fraction (4.1) is completely convergent for $y > 0$ if and only if the series

$$\sum_{p=1}^{\infty} (\beta_p + y)\,|\,B_{p-1}(z)\,|^2, \quad \text{where } z = x + iy \text{ and } \beta_p = \mathfrak{I}(b_p), \tag{4.6}$$

diverges. When (4.6) diverges, then the radius $r_n(z)$ of the circle $K_n(z)$ tends to 0 as n tends to ∞. The J-fraction is then convergent for this value of z, and its value is $f(z)$, the point common to all the circles $K_n(z)$. We shall prove that if the series (4.6) diverges for one value of $z = x + iy$, where $y > 0$, then it diverges for all such values of z. This will be accomplished by obtaining a condition for complete convergence which is independent of z, namely:

THEOREM 4.1. *The J-fraction (4.1) is completely convergent for $\mathfrak{I}(z) > 0$ if and only if at least one of the two infinite series*

$$\sum_{p=1}^{\infty} (1 + \beta_p)\,|\,B_{p-1}(0)\,|^2, \tag{4.7}$$

or

$$\sum_{p=1}^{\infty} (1 + \beta_p)\,|\,A_{p-1}(0)\,|^2 \tag{4.8}$$

is divergent.[10]

[10] See Hamburger [5] for the case b_p real.

It is easily seen that the series (4.6) and the series

$$\sum_{p=1}^{\infty} (\beta_p + y) | A_{p-1}(z) |^2 \tag{4.9}$$

converge or diverge together for $y > 0$. In fact, the ratio of corresponding terms is $| A_p(z)/B_p(z) |^2$, and there is the inequality

$$\frac{1}{| b_1 + z | + (a_1^2/y)} \leqq \left| \frac{A_p(z)}{B_p(z)} \right| \leqq \frac{1}{y}, \qquad y > 0. \tag{4.10}$$

The second part of the inequality is contained in (2.17). The first part follows then from the remark that $a_1^{-2}(b_1 + z - [B_n(z)/A_n(z)])$ is an approximant of another J-fraction obtained from (4.1) by advancing all indices by unity.

The determinant formula (2.10) shows that the polynomials $A_n(z)$, $B_n(z)$ do not vanish simultaneously. Therefore we conclude from (4.10) that neither vanishes in the upper half-plane $y > 0$. Now if z_1 is a zero of one of these polynomials and if $y \geqq 0$, the length of the vector $iy - z_1$ must increase as y increases. The same is true of any product of lengths of such vectors. Hence it follows that $| A_n(0) | < | A_n(iy) |$, $| B_n(0) | < | B_n(iy) |$ if $y > 0$. From these considerations we conclude that if at least one of the series (4.7), (4.8) diverges then the series $\sum_{p=1}^{\infty}(\beta_p + y) | B_{p-1}(iy) |^2$, i.e. the series (4.6) for $z = iy$, diverges for $y > 0$.

It remains to be shown that the series (4.6) diverges for any $z_0 = x_0 + iy_0$ with $y_0 > 0$. To do this, we show that any two points L_1, L_2 which are inside of every circle $K_n(z_0)$, $n = 1, 2, 3, \cdots$, are identical. In fact, we may select two sequences $\{u_n\}$ and $\{v_n\}$ lying in the lower half-plane such that

$$\lim_{n=\infty} t_1 t_2 \cdots t_n(z_0; u_n) = \lim_{n=\infty} \frac{A_n(z_0) - u_n A_{n-1}(z_0)}{B_n(z_0) - u_n B_{n-1}(z_0)} = L_1,$$

$$\lim_{n=\infty} t_1 t_2 \cdots t_n(z_0; v_n) = \lim_{n=\infty} \frac{A_n(z_0) - v_n A_{n-1}(z_0)}{B_n(z_0) - v_n B_{n-1}(z_0)} = L_2.$$

Let G be any bounded closed connected region in the upper half-plane $y > 0$ which contains on its interior the point z_0 and a portion of the positive half of the imaginary axis. The two sequences of rational functions of z,

$$\{t_1 t_2 \cdots t_n(z; u_n)\}, \qquad \{t_1 t_2 \cdots t_n(z; v_n)\} \tag{4.11}$$

are uniformly bounded over G. Hence we may select two subsequences, one from each, which are uniformly convergent over G to analytic limit-functions $f_1(z)$ and $f_2(z)$, respectively. Inasmuch as, for the pure imaginary points of G, $\lim r_n(iy) = 0$ so that $f_1(iy) = f_2(iy)$, it follows that $f_1(z) \equiv f_2(z)$ for all z in G. Therefore $L_1 = L_2$ and consequently (4.6) diverges if $\mathfrak{I}(z) = y > 0$.

We now suppose that both the series (4.7) and (4.8) converge, and shall prove that (4.6) converges for *all* values of z. The plan of the proof is as follows. We shall assign to u_n and v_n in (4.11) particular values in the lower half-plane in such a way that the sequences (4.11) will converge to limit-functions $f_1(z)$ and

$f_2(z)$ where $f_1(z) - f_2(z)$ is nowhere zero. This will of course imply that (4.6) converges for $\mathfrak{J}(z) > 0$. On account of the fact that none of the zeros of $B_p(z)$ are in the upper half-plane, we conclude by similar considerations as used before that (4.6) converges for all z.

On referring to (2.13) we recall that if $y \geqq 0$, then $\mathfrak{J}(B_n\bar{B}_{n-1}) \geqq 0$. Consequently, $\mathfrak{J}(\bar{B}_nB_{n-1}) \leqq 0$. Similarly, $\mathfrak{J}(\bar{A}_nA_{n-1}) \leqq 0$. Hence two sequences of constants lying in the lower half-plane are:

$$u_n = \frac{\bar{A}_n(0)}{\bar{A}_{n-1}(0)}, \qquad v_n = \frac{\bar{B}_n(0)}{\bar{B}_{n-1}(0)}, \qquad n = 1, 2, 3, \cdots.$$

These we shall use in forming the sequences (4.11). Put

$$(4.12)\quad \begin{array}{ll} \text{(i)} & P_n(z) = a_n(A_n(z)\bar{A}_{n-1}(0) - A_{n-1}(z)\bar{A}_n(0)), \\ \text{(ii)} & Q_n(z) = a_n(B_n(z)\bar{A}_{n-1}(0) - B_{n-1}(z)\bar{A}_n(0)), \\ \text{(iii)} & U_n(z) = a_n(A_n(z)\bar{B}_{n-1}(0) - A_{n-1}(z)\bar{B}_n(0)), \\ \text{(iv)} & V_n(z) = a_n(B_n(z)\bar{B}_{n-1}(0) - B_{n-1}(z)\bar{B}_n(0)), \end{array}$$

and (4.11) are then the sequences $\{P_n(z)/Q_n(z)\}$, $\{U_n(z)/V_n(z)\}$, respectively. Inasmuch as the numbers u_n and v_n lie in the lower half-plane, $P_n(z)/Q_n(z)$ and $U_n(z)/V_n(z)$ must lie in the circle $K_n(z)$. In view of the preceding remarks, the proof of Theorem 4.1 will be complete when we have proved the following theorem.

THEOREM 4.2. *If the series* (4.7) *and* (4.8) *both converge, then there exist four entire functions* $p(z)$, $q(z)$, $u(z)$, $v(z)$ *such that*

$$(4.13)\qquad p(z)v(z) - u(z)q(z) = 1,$$

and such that

$$(4.14)\quad \begin{array}{ll} \lim_{n=\infty} P_n(z) = p(z), & \lim_{n=\infty} Q_n(z) = q(z), \\ \lim_{n=\infty} U_n(z) = u(z), & \lim_{n=\infty} V_n(z) = v(z), \end{array}$$

uniformly over every bounded region of the z-plane.

PROOF.[11] From the determinant formula (2.10) we find by (4.12) that $P_n(z)V_n(z) - U_n(z)Q_n(z) = 1$, so that (4.13) will follow from (4.14). In the proof of (4.14) we shall use for the sake of brevity the notations:

$$\lambda_n(z) = (z + 2i\beta_{n+1})\bar{A}_n(0)^2$$

$$\mu_n(z) = (z + 2i\beta_{n+1})\bar{B}_n(0)^2$$

$$\nu_n(z) = (z + 2i\beta_{n+1})\bar{A}_n(0)\bar{B}_n(0).$$

[11] The essential idea of the proof is contained in several earlier investigations on continued fractions; cf., for instance, O. Perron [14], p. 235.

Since the series (4.7) and (4.8) converge and since $\beta_{n+1} \geqq 0$ it follows that the series

$$\sum \lambda_p(z), \qquad \sum \mu_p(z), \qquad \sum \nu_p(z),$$

and also the infinite product

$$\prod (1 - \nu_p(z))$$

converge absolutely and uniformly over any bounded closed region G of the z-plane. Moreover, there exists an N depending only upon G such that

$$\pi_n(z) = \prod_{p=N}^{N+n} (1 - \nu_p(z)) \neq 0, \qquad n = 0, 1, 2, \cdots, \quad z \text{ in } G.$$

On eliminating $A_{n+1}(z)$, $A_n(z)$, $A_{n-1}(z)$ from the recurrence formula (cf. (2.8))

$$a_{n+1}A_{n+1}(z) = (b_{n+1} + z)A_n(z) - a_nA_{n-1}(z),$$

and the equations (i) (written for n and $n + 1$) and (iii) of (4.12), we get a linear relation among $P_{n+1}(z)$, $U_n(z)$ and $P_n(z)$. Analogously we get all the following four identities

$$(4.15)\quad \begin{array}{ll} \text{(i)} & P_{n+1}(z) = \lambda_n(z)U_n(z) + (1 - \nu_n(z))P_n(z), \\ \text{(ii)} & Q_{n+1}(z) = \lambda_n(z)V_n(z) + (1 - \nu_n(z))Q_n(z), \\ \text{(iii)} & (1 - \nu_n(z))U_{n+1}(z) = -\mu_n(z)P_{n+1}(z) + U_n(z), \\ \text{(iv)} & (1 - \nu_n(z))V_{n+1}(z) = -\mu_n(z)Q_{n+1}(z) + V_n(z). \end{array}$$

If in (i) and (iii) we replace n by $N + n$ where N is the index introduced before, and if we use the notation

$$P_n^*(z) = \frac{P_{N+n}(z)}{\pi_{n-1}(z)}, \qquad U_n^*(z) = \pi_{n-1}(z)U_{N+n}(z),$$

$$b_n^*(z) = \frac{\lambda_{N+n}(z)}{\pi_{n-1}(z)\pi_n(z)}, \qquad c_n^*(z) = -\mu_{N+n}(z)\pi_{n-1}(z)\pi_n(z),$$

these relations become

$$(4.16)\quad \begin{array}{l} P_{n+1}^*(z) = b_n^*(z)U_n^*(z) + P_n^*(z), \\ U_{n+1}^*(z) = c_n^*(z)P_{n+1}^*(z) + U_n^*(z), \end{array} \qquad n = 1, 2, 3, \cdots.$$

By the remark at the beginning, the series $\sum b^*(z)$ and $\sum c^*(z)$ converge absolutely and uniformly over G, and hence there exists a finite number M_1 such that, for all z in G:

$$\prod_{p=1}^{\infty} (1 + | b_p^*(z) |) < M_1, \qquad \prod_{p=1}^{\infty} (1 + | c_p^*(z) |) < M_1.$$

Now, if $| U_1^*(z) | \leqq M_2$, $| P_1^*(z) | \leqq M_2$ over G, we have by (4.16):

$$| P_2^*(z) | \leqq | b_1^*(z) | \cdot | U_1^*(z) | + | P_1^*(z) | \leqq (1 + | b_1^*(z) |)M_2,$$

$$| U_2^*(z) | \leqq | c_1^*(z) | \cdot | P_2^*(z) | + | U_1^*(z) | \leqq (1 + | b_1^*(z) |)(1 + | c_1^*(z) |)M_2.$$

Continuing this procedure we find

$$| P_n^*(z) | \leqq M, \qquad | U_n^*(z) | \leqq M, \qquad n = 1, 2, 3, \cdots, z \text{ in } G,$$

where $M = M_2 M_1^2$.

Furthermore, by (4.16) we have $P_{n+1}^*(z) = P_1^*(z) + \sum_{p=1}^n b_p^*(z) U_p^*(z)$ and therefore

$$\lim_{n=\infty} P_{N+n+1}(z) = \lim_{n=\infty} \pi_n(z) P_{n+1}^*(z) = [\lim_{n=\infty} \pi_n(z)] \left[P_1^*(z) + \sum_{p=1}^{\infty} b_p^*(z) U_p^*(z) \right],$$

uniformly over G. This establishes the first limit in (4.14); the proof of the other limits can be made in the same way.

We have thus established in Theorem 4.1 a condition for the complete convergence of the J-fraction which does not depend upon z. This means, in fact, that when at least one of the constant term series (4.7) and (4.8) diverges, then the J-fraction (4.1) converges if $\mathfrak{J}(z) > 0$. Since the sequence of approximants is uniformly bounded for $\mathfrak{J}(z) \geqq k > 0$, the convergence is uniform over any finite region in this domain. We therefore have the following result.

THEOREM 4.3. *If at least one of the series* (4.7), (4.8) *diverges, the J-fraction* (4.1) *converges and represents an analytic function of z for $\mathfrak{J}(z) > 0$.*

We have also obtained in Theorem 4.2 the means which will enable us in the next section to answer completely the questions of convergence of the J-fraction and nature of the limit-function when both the series (4.7), (4.8) converge.

5. Simple convergence

On eliminating $A_{n-1}(z)$ from (i) and (iii) of (4.12) and $B_{n-1}(z)$ from (ii) and (iv), we now obtain the formula

$$\frac{A_n(z)}{B_n(z)} = \frac{P_n(z) - s_n U_n(z)}{Q_n(z) - s_n V_n(z)},$$

where $s_n = \bar{A}_n(0)/\bar{B}_n(0)$. When the series (4.7) and (4.8) both converge and $\lim_{n=\infty} s_n = s$ is finite, then the numerator and denominator converge to $p(z) - su(z)$ and $q(z) - sv(z)$, respectively. Since by (4.13) these two functions cannot vanish simultaneously, and $q(z) - sv(z)$ is not identically 0, being different from 0 for $\mathfrak{J}(z) > 0$, we conclude that the J-fraction converges to the quotient $[p(z) - su(z)]/[q(z) - sv(z)]$, which is a meromorphic function of z. The convergence is uniform in any closed bounded region containing no poles of the limit function. Similarly, if $\lim_{n=\infty} s_n = \infty$, we conclude that the J-fraction converges in like manner to the meromorphic function $u(z)/v(z)$. If the sequence $\{s_n\}$ has more than one limit-point (one of which may be ∞) it is easily seen with the aid of (4.13) that the J-fraction (4.1) is divergent for all values of z. These statements contain the following theorem.

THEOREM 5.1. *In case both the series* (4.7) *and* (4.8) *converge, then the convergence of the J-fraction* (4.1) *or of its reciprocal for a single value of z implies the convergence of the J-fraction or its reciprocal for any value of z. The value of the*

J-fraction is a meromorphic function of z, namely, in terms of the entire functions of (4.14),

$$\frac{p(z) - su(z)}{q(z) - sv(z)} \quad \text{if} \quad \lim_{n\to\infty} \frac{\bar{A}_n(0)}{\bar{B}_n(0)} = s \quad \text{is finite};$$

$$\frac{u(z)}{v(z)} \quad \text{if} \quad \lim_{n\to\infty} \frac{\bar{B}_n(0)}{\bar{A}_n(0)} = 0.$$

The J-fraction is not completely convergent in this case.

PART II

CONTINUED FRACTIONS AND LINEAR TRANSFORMATIONS IN INFINITELY MANY VARIABLES

6. J-matrices

In contrast with the preceding theory there is another theory in which the continued fraction appears in connection with a single linear transformation in infinitely many variables, namely:

$$\begin{aligned} (b_1 + z)x_1 - a_1x_2 \qquad\qquad\qquad\qquad &= y_1, \\ -a_1x_1 + (b_2 + z)x_2 - a_2x_3 \qquad\qquad &= y_2, \\ -a_2x_2 + (b_3 + z)x_3 - a_3x_4 &= y_3, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \end{aligned} \tag{6.1}$$

which carries the point $x = (x_1, x_2, x_3, \cdots)$ into the point $y = (y_1, y_2, y_3, \cdots)$. The matrix of this transformation is the *J-matrix* (4.5). For the present we allow the coefficients to be arbitrary complex numbers, the a_p being of course different from 0. If $y_1 = 1$, $y_n = 0$, $n > 1$, the equations (6.1) may be written as

$$x_1 = \frac{1}{b_1 + z - \dfrac{a_1x_2}{x_1}}, \quad \frac{x_2}{x_1} = \frac{a_1}{b_2 + z - \dfrac{a_2x_3}{x_2}}, \quad \frac{x_3}{x_2} = \frac{a_2}{b_3 + z - \dfrac{a_3x_4}{x_3}}, \quad \cdots, \tag{6.2}$$

and consequently x_1 is formally equal to the J-fraction:

$$\frac{1}{b_1 + z} - \frac{a_1^2}{b_2 + z} - \frac{a_2^2}{b_3 + z} - \cdots. \tag{6.3}$$

If the y_p are arbitrary, one could try to express a solution of (6.1) in the form

$$x_p = \sum_{q=1}^{\infty} \rho_{pq}y_q, \qquad p = 1, 2, 3, \cdots. \tag{6.4}$$

Here, ρ_{11} would be formally equal to the J-fraction (6.3). A matrix (ρ_{pq}) with this property is called a *right reciprocal* of the J-matrix (4.5). On substituting (6.4) into (6.1) we find the relations

$$-a_{p-1}\rho_{p-1,q} + (b_p + z)\rho_{p,q} - a_p\rho_{p+1,q} = \delta_{p,q}, \qquad p, q = 1, 2, 3, \cdots, \tag{6.5}$$

where a_0 is to be set equal to 0, and $\delta_{p,q}$ is equal to 0 or 1 according as $p \neq q$ or $p = q$. Since $a_p \neq 0$, $p = 1, 2, 3, \cdots$, it follows that, for a fixed q, if $\rho_{1,q}$ is chosen arbitrarily, then $\rho_{p,q}$, $p = 2, 3, 4, \cdots$, are uniquely determined. Therefore, there are infinitely many different right reciprocals.

An essential relationship to the J-fraction can be expected only for those reciprocals which belong to certain restricted classes. The most important class is given by the following definition:[12]

DEFINITION 6.1. *The matrix* (k_{pq}) *and the bilinear form in infinitely many variables*

$$K(\xi, \eta) = \sum_{p,q=1}^{\infty} k_{pq}\xi_p\eta_q$$

is said to be bounded if there exists a fixed number M *such that for all values of the* ξ_p *and* η_q *and for all integers* n

$$\left| \sum_{p,q=1}^{n} k_{pq}\xi_p\eta_q \right| \leqq M \cdot \left(\sum_{p=1}^{n} |\xi_p|^2 \right)^{1/2} \cdot \left(\sum_{q=1}^{n} |\eta_q|^2 \right)^{1/2}. \tag{6.6}$$

The explicit formulas for the $\rho_{p,q}$ in terms of the arbitrary $\rho_{1,q}$ and the polynomials $B_p(z)$ and $A_p(z)$ are:

$$\rho_{p,q} = \begin{cases} \rho_{1,q}B_{p-1}(z), & p = 1, 2, 3, \cdots, q; \\ \rho_{1,q}B_{p-1}(z) + A_{q-1}(z)B_{p-1}(z) - A_{p-1}(z)B_{q-1}(z), & \\ & p = q+1, q+2, \cdots. \end{cases} \tag{6.7}$$

This may be readily verified by comparing the recurrence formulas for the $A_p(z)$ and $B_p(z)$ with (6.5).

If one introduces new arbitrary functions $w_q(z)$ by means of the equations

$$\rho_{1,q} = B_{q-1}(z)w_q(z) - A_{q-1}(z), \qquad q = 1, 2, 3, \cdots,$$

then the formulas (6.7) take the form

$$\rho_{p,q}(z) = \begin{cases} B_{p-1}(z)B_{q-1}(z)\left(w_q(z) - \dfrac{A_{q-1}(z)}{B_{q-1}(z)}\right), & p = 1, 2, 3, \cdots, q; \\ B_{p-1}(z)B_{q-1}(z)\left(w_q(z) - \dfrac{A_{p-1}(z)}{B_{p-1}(z)}\right), & p = q+1, q+2, \cdots. \end{cases} \tag{6.8}$$

From (6.8) one sees immediately that the matrix (ρ_{pq}) is symmetric if and only if $w_1 = w_2 = w_3 = \cdots$; i.e. if and only if[13]

$$\rho_{n+1,q}/\rho_{n,q} = v_n, \qquad n \geqq q,$$

where v_n is independent of q, $(q, n = 1, 2, 3, \cdots)$.

---

[12] For the theory of bounded matrices see, for instance, Hellinger-Toeplitz [9], [10]. Cf. also H. T. Davis [3].

[13] See Hellinger [6]. Following the procedure which Weyl [18] had applied to boundary value problems of ordinary differential equations, Hellinger used a real parameter t and considered for $q = 1$ the equation $\rho_{n+1,q} = t\rho_{n,q}$ as a boundary condition. Beth [1] did the same for $q > 1$.

We shall suppose now, as in Part I, *that the* a_p *are real and positive and that the* b_p *satisfy the condition* $\mathfrak{I}(b_p) \geqq 0$, $p = 1, 2, 3, \cdots$. The next theorem gives a necessary and sufficient condition for the general right reciprocal of J to be bounded in the *limit-circle case* (cf. §2).

THEOREM 6.1. *The right reciprocal* (6.7) *is bounded in the limit-circle case if and only if the series*

$$\sum_{q=1}^{\infty} |\rho_{1,q}|^2 \tag{6.9}$$

is convergent.

PROOF. The necessity of the condition follows from the fact that in a bounded matrix the moduli of the elements of any row have a convergent sum of squares. If we recall that in the limit-circle case the series

$$\sum_{p=1}^{\infty} |A_{p-1}(z)|^2, \qquad \sum_{p=1}^{\infty} |B_{p-1}(z)|^2 \tag{6.10}$$

converge for all values of z, then we find from (6.7) that the convergence of (6.9) implies the convergence of the double series $\sum_{p,q=1}^{\infty} |\rho_{p,q}(z)|^2$. Consequently, the matrix $(\rho_{p,q})$ is not only bounded but is even completely continuous. We therefore have

COROLLARY 6.1. *In the limit-circle case, any right reciprocal* (6.7) *which is bounded is also completely continuous.*

It is possible for the series (6.10) to converge also in the limit-point case. Then there will be infinitely many bounded reciprocals. However, we can modify the condition of boundedness so that it is satisfied by only one such reciprocal. Definition 6.1 introduces an upper bound M for the values of the forms in (6.6), for $n = 1, 2, 3, \cdots$, as the variables run over the *spheres* $\sum_{p=1}^{n} |\xi_p|^2 \leqq 1$. Analogously, we can consider an upper bound as the variables run over the *ellipsoids* $\sum_{p=1}^{n} |\xi_p|^2(1 + \beta_p)^{-1} \leqq 1$, and accordingly we make the following definition:

DEFINITION 6.2. *The matrix* (ρ_{pq}) *will be said to be E-bounded if there exists a fixed number* M *such that for all values* ξ_p, η_q, *and for all integers* n:

$$\left| \sum_{p,q=1}^{n} \rho_{p,q}\xi_p\eta_q \right| \leqq M \cdot \left(\sum_{p=1}^{n} \frac{|\xi_p|^2}{1+\beta_p} \right)^{1/2} \cdot \left(\sum_{q=1}^{n} \frac{|\eta_q|^2}{1+\beta_q} \right)^{1/2}, \qquad \beta_p \geqq 0. \tag{6.11}$$

This is equivalent to saying that the matrix $((1 + \beta_p)^{1/2}(1 + \beta_q)^{1/2}\rho_{pq})$ is bounded in the sense of Definition 6.1, and hence the theorems about bounded matrices can be extended immediately to E-bounded matrices.

THEOREM 6.2. *In the limit-point case, let* $f(z)$ *be the analytic function whose value for every* z *with* $\mathfrak{I}(z) > 0$ *is common to all the circles* $K_n(z)$. *Then the formula* (6.8) *with* $w_q(z) = f(z)$, $q = 1, 2, 3, \cdots$, *gives the unique E-bounded right reciprocal, which is simultaneously the unique left reciprocal.*

PROOF. We show first that if $w_q(z) = f(z)$ for all q and for $\mathfrak{J}(z) > 0$, then the matrix (ρ_{pq}) is E-bounded. If we suppose only $w_1(z) = w_2(z) = \cdots = w_n(z)$, where n is a fixed index, then from (6.8) it follows that

$$(6.12) \qquad \frac{\rho_{n+1,q}}{\rho_{n,q}} = \frac{A_n(z) - w_1(z)B_n(z)}{A_{n-1}(z) - w_1(z)B_{n-1}(z)}, \qquad q = 1, 2, 3, \cdots, n.$$

Hence, the quotient $\rho_{n+1,q}/\rho_{n,q} = v_n$ is independent of q:

$$(6.13) \qquad \rho_{n+1,q} = v_n\rho_{n,q}, \qquad q = 1, 2, 3, \cdots, n.$$

If we identify (6.12) with (2.11), we conclude at once that $\mathfrak{J}(v_n) \leqq 0$ if and only if $w_1(z)$ has its value in the circle $K_n(z)$.

Let $\xi_1, \xi_2, \cdots, \xi_n$ be arbitrary real numbers. On multiplying both members of (6.13) by ξ_q and summing over q from 1 to n we then have:

$$y_{n+1} = v_n y_n, \qquad y_p = \sum_{q=1}^{n} \rho_{pq}\xi_q .$$

Therefore $y_{n+1}\bar{y}_n = v_n \mid y_n \mid^2$ and consequently

$$(6.14) \qquad \mathfrak{J}(y_{n+1}\bar{y}_n) \leqq 0$$

if $w_1(z)$ is in $K_n(z)$.

We now multiply the equation (6.5) by ξ_q and sum over q from 1 to n. This gives the equation

$$(6.15) \quad L_p(y) \equiv -a_{p-1}y_{p-1} + (b_p + z)y_p - a_py_{p+1} = \xi_p, \qquad p = 1, 2, 3, \cdots, \quad a_0 = 0.$$

Now, one has immediately the identity[14]

$$\sum_{p=1}^{n} (\bar{y}_pL_p(y) - y_1\overline{L_p(y)}) = 2i\left(\sum_{p=1}^{n} (\beta_p + \mathfrak{J}(z))| y_p |^2 - a_n\mathfrak{J}(y_{n+1}\bar{y}_n)\right)$$

and by (6.15) we get

$$(6.16) \qquad a_n\mathfrak{J}(y_{n+1}\bar{y}_n) = \sum_{p=1}^{n} (\beta_p + \mathfrak{J}(z))| y_p |^2 + \sum_{p=1}^{n} \xi_p\mathfrak{J}(y_p).$$

Considering now the quadratic form

$$R_n(\xi, \xi) = \sum_{p,q=1}^{n} \rho_{pq}\xi_p\xi_q = \sum_{p=1}^{n} y_p\xi_p,$$

---

[14] This identity is analogous to one of the so-called "Green's formulas" in the theory of differential equations. It may be emphasized that if we regard $\rho_{11} = w$ as a complex variable, then (6.16) along with (6.14) gives the inequality defining our nest of circular regions. This is the method which was used by Weyl [18] for differential equations, and by Hellinger [6] and Beth [1] for real J-fractions. The method has been used also for other problems, cf., for instance, Weyl [19].

we have by Schwarz's inequality

$$|R_n(\xi, \xi)|^2 \leqq \sum_{p=1}^{n} \frac{\xi_p^2}{\beta_p + \Im(z)} \cdot \sum_{p=1}^{n} (\beta_p + \Im(z))|y_p|^2,$$

and, using (6.16) and (6.14):

$$|R_n(\xi, \xi)|^2 \leqq \sum_{p=1}^{n} \frac{\xi_p^2}{\beta_p + \Im(z)} \cdot \left(-\sum_{p=1}^{n} \xi_p \Im(y_p)\right).$$

Inasmuch as $-\sum_{p=1}^{n} \xi_p \Im(y_p) = -\Im(R_n(\xi, \xi)) \leqq |R_n(\xi, \xi)|$, we have: $|R_n(\xi, \xi)| \leqq \sum_{p=1}^{n} \xi_p^2/(\beta_p + \Im(z))$, or

$$\left|\sum_{p,q=1}^{n} \rho_{pq}\xi_p\xi_q\right| \leqq \sum_{p=1}^{n} \frac{\xi_p^2}{\beta_p + \Im(z)} \leqq M_1 \cdot \sum_{p=1}^{n} \frac{\xi_p^2}{1 + \beta_p}, \tag{6.17}$$

for a suitably chosen M_1 depending only upon the domain of z. For the related bilinear form we therefore have

$$\begin{aligned} |R_n(\xi, \eta)| &= \tfrac{1}{4}|R_n(\xi + \eta, \xi + \eta) - R_n(\xi - \eta, \xi - \eta)| \\ &\leqq \frac{1}{2} \sum_{p=1}^{n} \frac{\xi_p^2 + \eta_p^2}{\beta_p + \Im(\xi)}. \end{aligned} \tag{6.18}$$

It follows that if $f(z)$ is the value common to all the circles $K_n(z)$, i.e. the value $\lim (A_p(z)/B_p(z))$ of the J-fraction, and if every $w_q(z) = f(z)$, then the matrix (ρ_{pq}) is E-bounded.

It remains to be seen that any other reciprocal given by (6.8) is not E-bounded in the limit-point case. In fact, for any other reciprocal there is at least one z and integers q and $N \geqq q$ such that

$$\left|w_q(z) - \frac{A_{p-1}(z)}{B_{p-1}(z)}\right| \geqq d > 0 \quad \text{for} \quad p \geqq N.$$

Hence we see immediately by the second equation (6.8) that $\sum_p (1 + \beta_p) \cdot (1 + \beta_q) |\rho_{pq}|^2$ diverges inasmuch as $\sum_p (1 + \beta_p) |B_{p-1}(z)|^2$ diverges. This implies that the matrix $((1 + \beta_p)^{1/2}(1 + \beta_q)^{1/2}\rho_{pq})$ is not bounded, i.e. the matrix (ρ_{pq}) is not E-bounded.

From the inequality (6.18) we shall now derive some estimates for $\rho_{pq} = \rho_{pq}(z)$ which will be useful later on.

THEOREM 6.3. *Let $w_q(z) = f(z)$, $q = 1, 2, 3, \cdots, n$, in (6.8), where $f(z)$ is in the circle $K_n(z)$ for $\Im(z) > 0$. Then*

$$|\rho_{pq}(z)| \leqq \frac{1}{(y + \beta_p)^{1/2}(y + \beta_q)^{1/2}} \leqq \frac{1}{y}, \tag{6.19}$$

$$z = x + iy, \qquad y > 0, \qquad p, q \leqq n;$$

and

$$\rho_{pq}(z) = \frac{\delta_{pq}}{z} + \frac{\theta_{pq}(z)}{zy} G_p, \tag{6.20}$$

$$|\theta_{pq}| \leqq 1, \qquad G_p = a_p + |b_p| + a_{p-1}, \qquad p, q < n.$$

PROOF. The inequality (6.19) is an immediate consequence of (6.18) if we specialize the variables such that one $\xi_p^2 = \Im(z) + \beta_p$, one $\eta_q^2 = \Im(z) + \beta_q$, and the other variables are 0. Now by (6.5):

$$\rho_{pq}(z) = \frac{\delta_{pq}}{z} + \frac{a_p \rho_{p+1,q}(z) - b_p \rho_{pq}(z) + a_{p-1} \rho_{p-1,q}(z)}{z}.$$

Applying (6.19) to this identity we obtain (6.20).

7. A general theorem of invariability

We have seen that the behavior of the J-fraction (4.1) is invariant in some respect under change in the particular value of z in the upper half-plane. This has been implied by the fact that if the series (4.6) converges for one z with $\Im(z) > 0$, it converges for all z. We now derive a more general theorem of invariability which covers *entirely arbitrary* J-fractions. The proof reveals in a certain way the inner structure of the theory.

THEOREM 7.1. *Let* (4.1) *be a J-fraction with entirely arbitrary coefficients $a_p \neq 0$ and b_p. Let α_p, $p = 1, 2, 3, \cdots$, be arbitrary real numbers not less than* 1. *Then, if the two series*

$$\sum_{p=1}^{\infty} \alpha_p^2 \,|\, A_{p-1}(z) \,|^2, \qquad \sum_{p=1}^{\infty} \alpha_p^2 \,|\, B_{p-1}(z) \,|^2 \tag{7.1}$$

converge simultaneously for one value of z, they converge for every value of z.

This obviously contains the theorem of invariability of §4. Moreover, it supplements the remark on the limit-point case after Corollary 6.1 in this way: *If the series* (6.10) *both converge for one value of z, then there exist infinitely many bounded reciprocals for every value of z.*

PROOF OF THEOREM 7.1.[15] Write, as in (6.15) but with a_0 now equal to 1,

$$L_p(y) \equiv -a_{p-1}y_{p-1} + (b_p + z)y_p - a_p y_{p+1}, \quad p = 1, 2, 3, \cdots, a_0 = 1,$$

and denote by $L_p^*(y)$ the same expression with z replaced by z^*. The solution of the system $L_p(y) = 0$ under the initial conditions $y_0 = -1$, $y_1 = 0$ is $y_p = A_{p-1}(z) = A_{p-1}$, while under the initial conditions $y_0 = 0$, $y_1 = 1$ the solution is $y_p = B_{p-1}(z) = B_{p-1}$. If y_p, y_p^* are arbitrary solutions of the systems $L_p(y) = 0$ and $L_p^*(y) = 0$, respectively, then we obtain immediately the relation:

$$\sum_{p=1}^{n} (y_p^* L_p(y) - y_p L_p^*(y^*)) = y_1 y_0^* - y_0 y_1^*$$
$$- a_n(y_{n+1} y_n^* - y_n y_{n+1}^*) + (z - z^*) \sum_{p=1}^{n} y_p y_p^* = 0.$$

---

[15] Cf. footnote 4. This proof uses the idea which Weyl [19] has applied in similar problems, namely, to express the relationship between solutions for two different parameter values as a Volterra integral or sum equation. This procedure as well as the procedure used by Weyl [18] and Hellinger [6] may be embraced in a general set-up, if one uses an arbitrary one of the infinitely many reciprocals (6.7) of the J-matrix. Then, the different forms of the proof appear in specializing the reciprocal in different ways, for instance so that it becomes symmetrical ($\rho_{pq} = \rho_{qp}$) or a Volterra form ($\rho_{pq} = 0$ for $p < q$).

In particular, for $y_p = A_{p-1}$ and $y_p = B_{p-1}$ we get

$$y_1^* - a_n(y_n^* A_n - y_{n+1}^* A_{n-1}) + (z - z^*) \sum_{p=1}^{n} y_p^* A_{p-1} = 0,$$

$$y_0^* - a_n(y_n^* B_n - y_{n+1}^* B_{n-1}) + (z - z^*) \sum_{p=1}^{n} y_p^* B_{p-1} = 0,$$

respectively. On multiplying the first of these equations by $-B_{n-1}$, the second by A_{n-1}, and then adding, we get:

$$y_n^* + (z^* - z) \sum_{p=1}^{n-1} (A_{p-1} B_{n-1} - A_{n-1} B_{p-1}) y_p^* = y_1^* B_{n-1} - y_0^* A_{n-1} .$$

Therefore, $\zeta_p = \alpha_p y_p^*$ is a solution of the Volterra sum equation:

$$\zeta_p + \sum_{q=1}^{p-1} k_{pq} \zeta_q = g_p , \qquad p = 1, 2, 3, \cdots , \tag{7.2}$$

in which

$$k_{pq} = \alpha_p \alpha_q^{-1} (z^* - z)(A_{q-1} B_{p-1} - A_{p-1} B_{q-1}), \qquad g_p = \alpha_p (y_1^* B_{p-1} - y_0^* A_{p-1}).$$

The proof of the theorem will be complete if we show that $\sum | \zeta_p |^2$ is convergent.

From the convergence of the series (7.1) it follows at once that g_p satisfies the condition:

$$C^2 = \sum_{p=1}^{\infty} | g_p |^2$$

is finite; and that, inasmuch as $\alpha_p \geqq 1$, the double series $\sum | k_{pq} |^2$ converges, so that for r sufficiently large:

$$\epsilon_r = \sum_{q=1}^{\infty} \sum_{p=r}^{\infty} | k_{pq} |^2 < 1. \tag{7.3}$$

We now multiply (7.2) by $\bar{\zeta}_p$ and sum over p from r to m, $m > r$. This gives, if we apply Schwarz's inequality:

$$\begin{aligned} \sum_{p=r}^{m} | \zeta_p |^2 &\leqq C \cdot \left(\sum_{p=r}^{m} | \zeta_p |^2 \right)^{1/2} + \sum_{p=r}^{m} \sum_{q=1}^{m} | k_{pq} \zeta_p \zeta_q | \\ &\leqq \left(\sum_{p=r}^{m} | \zeta_p |^2 \right)^{1/2} \left(C + \sum_{q=1}^{m} | \zeta_q | \left(\sum_{p=r}^{m} | k_{pq} |^2 \right)^{1/2} \right), \end{aligned}$$

and consequently, again using Schwarz's inequality,

$$\left(\sum_{p=r}^{m} | \zeta_p |^2 \right)^{1/2} \leqq C + \epsilon_r \left(\sum_{q=1}^{m} | \zeta_q |^2 \right)^{1/2} \leqq C + \epsilon_r \left(\sum_{q=1}^{r-1} | \zeta_q |^2 \right)^{1/2} + \epsilon_r \left(\sum_{q=r}^{m} | \zeta_q |^2 \right)^{1/2},$$

or

$$\left(\sum_{p=r}^{m} | \zeta_p |^2 \right)^{1/2} \cdot (1 - \epsilon_r) \leqq C + \epsilon_r \left(\sum_{q=1}^{r-1} | \zeta_q |^2 \right)^{1/2} .$$

Hence, by (7.3), the series $\sum | \zeta_p |^2$ converges, and the theorem is proved.

8. Asymptotic and integral expressions for the J-fraction

We shall now consider a different approach to the problem of relating to the J-fraction the leading element $\rho_{11}(z)$ of a reciprocal of the J-matrix (4.5). We suppose still that the J-fraction satisfies the conditions (4.2), and shall use the ideas of §6, particularly Theorem 6.3, to obtain conditions under which the approximants $A_n(z)/B_n(z)$ approximate to $\rho_{11}(z)$ in the asymptotic sense of the following definition.

DEFINITION 8.1. *Consider the domain*

$$S: \qquad \alpha \leqq \arg z \leqq \pi - \alpha \; (0 < \alpha < \pi/2), \qquad \mathfrak{J}(z) \geqq \delta > 0,$$

where α and δ are arbitrary positive numbers. A function $f(z)$ is said to be represented asymptotically by the J-fraction (4.1) *if*

$$\lim_{z=\infty} z^{2n}\left(f(z) - \frac{A_n(z)}{B_n(z)}\right) = 0, \qquad n = 1, 2, 3, \cdots, \tag{8.1}$$

as z approaches ∞ in S.

This is the same thing as to say that $f(z)$ is represented asymptotically by the series expansion of the J-fraction into powers of $1/z$.

THEOREM 8.1. *A function $f(z)$ is represented asymptotically by the J-fraction* (4.1) *if and only if for every n there exists a number M_n such that the value of $f(z)$ is in the circle $K_n(z)$ for z in S and $\mathfrak{J}(z) > M_n$.*[16]

PROOF. To prove that the condition is sufficient, let $w_q(z) = f(z)$, $q = 1, 2, 3, \cdots$, in (6.8), where $f(z)$ is any function satisfying the condition of the theorem. Then, from (6.8) we have:

$$z^{2p}\left(f(z) - \frac{A_p(z)}{B_p(z)}\right) = \frac{z^{2p}\rho_{p+1,p}(z)}{B_p(z)B_{p-1}(z)}, \qquad p = 1, 2, 3, \cdots. \tag{8.2}$$

If then $\mathfrak{J}(z) > M_{p+1}$, z in S, so that $f(z)$ has its value in $K_{p+1}(z)$, it follows from the formula (6.19) that

$$z^{2p}\left(f(z) - \frac{A_p(z)}{B_p(z)}\right) = \frac{H_p(z)}{\mathfrak{J}(z)},$$

where $H_p(z)$ is bounded in S; hence (8.1) holds.

We now suppose conversely that $f(z)$ is represented asymptotically by the J-fraction, and form the expression:

$$z^{2n+1}\left[\left(\frac{A_{n+1}(z)}{B_{n+1}(z)} - f(z)\right) + \left(f(z) - \frac{A_n(z)}{B_n(z)}\right)\right].$$

By the determinant formula, this is equal to $(a_1a_2 \cdots a_n)^2 + Q/z$, where Q is bounded in S. Hence we see that

$$\lim_{z=\infty} z^{2n+1}\left(f(z) - \frac{A_n(z)}{B_n(z)}\right) = (a_1a_2 \cdots a_n)^2, \tag{8.3}$$

[16] For the case where the b_p are real and $f(z)$ is analytic, R. Nevanlinna [12] has proved the same theorem with the stronger formulation that $M_n = 0$.

as z tends to ∞ in S. Let v_n be determined by the relation $f(z) = t_1 t_2 \cdots t_n(z; v_n)$; and recall that $f(z)$ is in the circle $K_n(z)$ if and only if $\mathfrak{J}(v_n) \leqq 0$. On substituting this value of $f(z)$ in (8.3) we obtain (cf. (2.7)):

$$\lim_{z=\infty} z^{2n+1}\left(\frac{A_n(z) - v_n A_{n-1}(z)}{B_n(z) - v_n B_{n-1}(z)} - \frac{A_n(z)}{B_n(z)}\right) = \lim_{z=\infty} z^{2n+1} \frac{v_n}{a_n B_n(z)[B_n(z) - v_n B_{n-1}(z)]}$$
$$= (a_1 a_2 \cdots a_n)^2.$$

Thus,

$$z v_n = \frac{z a_n B_n(z)^2[(a_1 a_2 \cdots a_n)^2 + \epsilon(z)]}{z^{2n+1} + a_n B_n(z) B_{n-1}(z)[(a_1 a_2 \cdots a_n)^2 + \epsilon(z)]},$$

where $\epsilon(z)$ approaches 0 as z approaches ∞ in S. Therefore, since the coefficient of z^n in $B_n(z)$ is $(a_1 a_2 \cdots a_n)^{-1}$, $z v_n$ converges to the positive limit a_n. Consequently, there exists a number M_n such that $\mathfrak{J}(v_n) \leqq 0$ and $f(z)$ is in $K_n(z)$. if $\mathfrak{J}(z) > M_n$ and z is in S.

Furthermore, Theorem 6.3 enables us to connect with the J-fractions "Stieltjes integrals" of the form (1.2) with range of integration extended over the whole real axis. Any function $f(z)$ which, for $\mathfrak{J}(z) > 0$, is analytic and has its values in the circles $K_n(z)$, $n = 1, 2, 3, \cdots$, will be called *equivalent* to the J-fraction. It is to be recalled that in the limit-point case, there is but one equivalent function, namely the value of the J-fraction; while in the limit-circle case there are infinitely many. If in (6.8) we suppose that $w_q(z) = f(z)$, $q = 1, 2, 3, \cdots$, then from (6.20) with $p = q = 1$ we obtain for any equivalent function the following estimate:

$$f(z) = \frac{1}{z} + \frac{g(z)}{z\mathfrak{J}(z)}, \quad \text{where} \quad |g(z)| \leqq C \quad \text{if} \quad \mathfrak{J}(z) > 0, \tag{8.4}$$

C being independent of z. From this we deduce the following general theorem:

THEOREM 8.2. *A function $f(z)$ has a Stieltjes integral representation of the form*

$$f(z) = \int_{-\infty}^{+\infty} \frac{d\phi(u)}{z + u}, \qquad \phi(+\infty) - \phi(-\infty) = 1, \tag{8.5}$$

in which $\phi(u)$ is real, bounded, and monotone nondecreasing, if $f(z)$ satisfies all of the following three conditions for $\mathfrak{J}(z) > 0$:

(i) *$f(z)$ is analytic*; (ii) $\mathfrak{J}[f(z)] \leqq 0$; (iii) The estimate (8.4) *holds*.*

PROOF. Suppose that $0 < y < c$, and consider the contour Γ in the upper half of the z-plane consisting of: the straight line segment from $A = -c^2 + iy$ to $B = c^2 + iy$, the straight line segment from B to $B' = c^2 + ic$, the arc of the circle with center at the origin through B' to $A' = -c^2 + ic$, and, finally, the

---

* The condition (iii) is not necessary, as can be seen from the example $\phi(u) = 0$ for $u \leqq 1, \phi(u) = 1 - u^{-(1/2)}$ for $u > 1$. This shows simultaneously that not every integral of the form (8.5) is equivalent to a J-fraction.

straight line segment from A' to A. Inasmuch as $f(z)$ is analytic in the domain interior to Γ we have, using (8.4), and evaluating explicitly $\int_{A'}^{B'} dz/z$:

$$\int_A^B f(z)\,dz = -\int_B^{B'} f(z)\,dz - \int_{B'}^{A'} f(z)\,dz - \int_{A'}^{A} f(z)\,dz$$

$$= -\pi i + 2i \arctan (y/c^2) + \frac{H}{c^2}\log\frac{y}{c} + \frac{H_1}{c},$$

where H and H_1 are bounded as c tends to ∞. Hence it follows that

$$\lim_{c=\infty}\int_{-c^2}^{+c^2} f(x+iy)\,dx = -\pi i, \quad \text{if} \quad y > 0. \tag{8.6}$$

Let $f(x+iy) = v(x, y) - iw(x, y)$, where $v(x, y)$ and $w(x, y)$ are real functions; then by (ii) $w(x, y) \geqq 0$ for $y > 0$. From (8.6) we conclude that

$$\lim_{c=\infty}\int_{-c^2}^{+c^2} w(x, y)\,dx = \pi. \tag{8.7}$$

Moreover, $\psi(u, y) = \int_0^u w(x, y)\,dx$ is a monotone nondecreasing function of u, is bounded by (8.7), and

$$\psi(+\infty, y) - \psi(-\infty, y) = \pi. \tag{8.8}$$

A well-known theorem[17] states that there exists a bounded monotone nondecreasing function $\psi(u)$ such that $\psi(+\infty) - \psi(-\infty) = \pi$, and a sequence $y_1, y_2, y_3, \cdots$ of positive numbers approaching the limit 0 such that

$$\lim_{n=\infty} \psi(u, y_n) = \psi(u) \tag{8.9}$$

at all points u where $\psi(u)$ is continuous.

If z is any point within Γ, then Cauchy's integral formula gives:

$$f(z) = \frac{1}{2\pi i}\int_\Gamma \frac{f(s)\,ds}{s - z}.$$

Using (8.4) one may readily verify that for $c \to \infty$ this goes over into

$$f(z) = \frac{1}{2\pi i}\int_{-\infty}^{+\infty} \frac{f(u+iy)\,du}{u + iy - z}, \tag{8.10}$$

where the integral is to be regarded in the sense of Cauchy's principal value. Let z^* be the point outside Γ which is symmetrical to the point z with respect to the line segment AB. Then we must have:

$$0 = \frac{1}{2\pi i}\int_{-\infty}^{+\infty} \frac{f(u+iy)\,du}{u + iy - z^*} \quad \text{or} \quad 0 = \frac{1}{2\pi i}\int_{-\infty}^{+\infty} \frac{\overline{f(u+iy)}\,du}{u - iy - \bar{z}^*}.$$

[17] For the proof see, for instance, Perron [14], pp. 394–395. This theorem has been applied in almost all investigations on problems of the kind considered here. The idea goes back to Stieltjes, and was developed and extended by Hilbert as one of the most important tools in his theory of infinite quadratic forms. See Hilbert [11] (book), p. 113 and 116.

Inasmuch as $u + iy - z = u - iy - \bar{z}^*$ we then have, on subtracting the last equation from the equation (8.10) and then introducing the function $\psi(u, y)$:

$$f(z) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{-w(u, y)\, du}{u + iy - z} = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{d_u \psi(u, y)}{z - u - iy}.$$

On letting y approach 0 over the sequence y_n for which (8.9) holds, one then finds by a well-known argument (see footnote 17):

$$f(z) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{d\psi(u)}{z - u},$$

or, if we define $\phi(u)$ by $\phi(u) = -(1/\pi)\psi(-u)$, this becomes (8.5). Since $\phi(u)$ is bounded and monotone, the integral converges absolutely (not merely as Cauchy's principal value) and uniformly for z in any region at a positive distance from the real axis. The function $\phi(u)$ is given at all points of continuity by

$$(8.11) \qquad \pi \cdot \phi(u) = \lim_{y=0} \int_0^{-u} \Im[f(x + iy)]\, dx,$$

where y approaches 0 over the sequence y_n. Since $f(z)$ is now expressed as an integral (8.5) the inversion process of Stieltjes[18] gives $\phi(u)$ in terms of $f(z)$, and shows, simultaneously, that (8.11) holds no matter how y approaches 0 over positive values. Thus $\phi(u)$ is determined uniquely by $f(z)$ to an additive constant at all points of continuity.

That the integrals ("moments")

$$(8.12) \qquad \int_{-\infty}^{+\infty} u^p d\phi(u), \qquad p = 0, 1, 2, \cdots,$$

do not all exist when a single one of the b_p is nonreal may be argued from the fact that they are real if they exist, and from a theorem of H. Hamburger.[19]

The considerations of this section are closely connected with the "moment problem": To determine a real bounded nondecreasing function $\phi(u)$ taking on infinitely many different values and satisfying the infinite system of equations:

$$(8.13) \qquad c_p = \int_{-\infty}^{+\infty} u^p d\phi(u), \qquad p = 0, 1, 2, \cdots,$$

where the c_p are given real numbers.

By the theorem of Hamburger just cited, $\phi(u)$ is a solution of (8.13) if and only if the function

$$(8.14) \qquad f(z) = \int_{-\infty}^{+\infty} \frac{d\phi(u)}{z + u}$$

---

[18] Stieltjes [15], Chapter VI.

[19] Hamburger [5], Part I, Theorem IX (p. 268 ff).

is represented asymptotically by the power series $P(1/z)$ with the given numbers as coefficients:

$$f(z) \sim \frac{c_0}{z} - \frac{c_1}{z^2} + \frac{c_2}{z^3} - \cdots . \tag{8.15}$$

This is the same as saying that $f(z)$ is represented asymptotically by the J-fraction (4.1) "associated" (cf. Perron [14], §61) with $P(1/z)$. Hence, *all* the solutions of the moment problem are obtained by finding all functions $f(z)$ of the form (8.14) asymptotically equal to the J-fraction. By the theorem of Nevanlinna mentioned above and Theorem 8.2 these functions $f(z)$ are *just* those functions "equivalent" to the J-fraction. In this way one arrives at the complete solution of the moment problem.

It follows immediately that the moment problem is "determinate" in the limit-point case, and "indeterminate" in the limit-circle case.

Northwestern University,
Evanston, Illinois.

Bibliography

1. Richard A. Beth, *Untersuchungen über die Spektraldarstellung von J-Formen*, Thesis, University of Frankfurt a.M., 1932, 69 pages.
2. T. Carleman, *Sur les équations intégrales singulières à noyau symétrique*, Uppsala, 1923, pp. 189–220.
3. H. T. Davis, *The Theory of Linear Operators*, Bloomington, 1936, Chapters III and XII.
4. J. Grommer, *Ganze transcendente Funktionen mit lauter reellen Nullstellen*, Journal für die reine und angewandte Mathematik, vol. 144 (1914), pp. 212–238
5. H. Hamburger, *Über eine Erweiterung des Stieltjesschen Momentenproblems*, Parts I, II, and III, Mathematische Annalen, vol. 81 (1920), pp. 235–319; vol. 82 (1920) pp. 120–164; and pp. 168–187.
6. E. Hellinger, *Zur Stieltjesschen Kettenbruchtheorie*, Mathematische Annalen, vol. 86 (1922), pp. 18–29.
7. ———, *Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen*, Journal für die reine und angewandte Mathematik, vol. 136 (1909), pp. 210–271.
8. E. Hellinger and O. Toeplitz, *Zur Einordnung der Kettenbruchtheorie in die Theorie der quadratischen Formen von unendlichvielen Veränderlichen*, Journal für die reine und angewandte Mathematik, vol. 144 (1914), pp. 212–238.
9. ———, *Grundlagen für eine Theorie der unendlichen Matrizen*, Mathematische Annalen, vol. 69 (1910), pp. 289–330.
10. ———, *Integralgleichungen und Gleichungen mit unendlichvielen Unbekannten*, Encyklopädie der mathematischen Wissenschaften, II C 13, No. 18 and No. 43.
11. D. Hilbert, *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen*, Vierte Mitteil., Nachrichten der königlichen Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse, 1906, p. 219 ff; or book of same title, Leipzig, 1912, Chapter IV.
12. R. Nevanlinna, *Asymptotische Entwicklungen beschränkter Funktionen und das Stieltjessche Momentenproblem*, Annales academiae Fennicae, Series A, vol. 18 (1922), No. 5.
13. J. F. Paydon and H. S. Wall, *The continued fraction as a sequence of linear transformations*, Duke Mathematical Journal, vol. 9 (1942), pp. 360–372.

14. O. PERRON, *Die Lehre von den Kettenbrüchen*, 2nd ed., Leipzig and Berlin, 1929.
15. T. J. STIELTJES, *Recherches sur les fractions continues*, Oeuvres, vol. II, pp. 402–566. Published also in vols. 8 and 9 of Annales de Toulouse; and in vol. 32 of Mémoires présentés à l'Académie des Sciences de l'Institut National de France.
16. E. B. VAN VLECK, *On an extension of the 1894 memoir of Stieltjes*, Transactions of the American Mathematical Society, vol. 4 (1903), pp. 297–332.
17. ———, *On the convergence of continued fractions with complex elements*, Transactions of the American Mathematical Society, vol. 2 (1901), pp. 215–233.
18. H. WEYL, *Über gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen*, Mathematische Annalen, vol. 68 (1910), pp. 220–269.
19. ———, *Über das Pick-Nevanlinna'sche Interpolationsproblem und sein infinitesimales Analogon*, Annals of Mathematics, Series 2, vol. 36 (1935), pp. 230–254.

REMARKS ON TWO-LEAVED ORIENTABLE COVERING MANIFOLDS OF CLOSED MANIFOLDS

BY TSAI-HAN KIANG

(Received October 5, 1942)

1. In the present note the covering complexes considered are *without branch point*, and the covering manifolds are, in addition, *of finite number of leaves.*[1] Our purpose is to establish the following two theorems.

THEOREM 1. *Any orientable covering manifold of a closed non-orientable n-manifold M is a covering manifold of the 2-leaved orientable covering manifold*[2] *of M, and hence a covering manifold of M of even number of leaves.*

THEOREM 2. *A necessary and sufficient condition for the existence of a simplicial topological self-mapping of a closed orientable n-manifold $\bar{M}$, which is involutory, without fixed point and orientation-preserving (orientation-reversing),*[3] *is that $\bar{M}$ be a 2-leaved covering manifold of a closed orientable (nonorientable) n-manifold M.*

2. Theorem 1 follows at once from the following two lemmas.

LEMMA 1. *Let K be a connected n-complex and F its fundamental group. Suppose that G and H are two subgroups of F and that H is a subgroup of G. If K^* and K^{**} are the covering complexes of K, determined by G and H respectively,*[4] *then K^{**} is a covering complex of K^*.*[5]

PROOF. From hypothesis the fundamental groups F^* and F^{**} of K^* and K^{**} are simply isomorphic to G and H respectively.[4] By virtue of the simple isomorphism between F^* and G, there is a definite subgroup of F^*, which is simply isomorphic to F^{**}. Let $\bar{K}$ be the covering complex of K^*, determined by this subgroup of the fundamental group F^* of K^*. Then $\bar{K}$ is a covering complex of K.[6] Since the fundamental group of $\bar{K}$ is simply isomorphic to F^{**} and therefore to H, and since a covering complex of K is uniquely determined by a subgroup of the fundamental group of K,[4] $\bar{K}$ and K^{**} are not distinct.[7] Hence K^{**} is a covering complex of K^*.

LEMMA 2. *Let $\bar{M}$ be a covering manifold*[8] *of a closed n-manifold M with a cellular decomposition M_a,*[9] *and $\bar{M}_a$ the cellular decomposition of $\bar{M}$ derived from*

---

[1] The topological terms will be defined as in Seifert-Threlfall, *Lehrbuch der Topologie* (1934). This book will be referred to hereafter as ST.

[2] Cf. ST, p. 272.

[3] ST, p. 129.

[4] ST, p. 193.

[5] Cf. K. Reidemeister, *Einführung in die Kombinatorische Topologie* (1932), p. 125.

[6] ST, p. 194.

[7] ST, pp. 182–183.

[8] A finite covering complex of a closed n-manifold is obviously a closed n-manifold.

[9] ST, p. 341.

M_a.[10] *If an edge-path $\bar{U}$ on $\bar{M}_a$ at a vertex $\bar{O}_1$ is closed, then $\bar{U}$ and its closed ground-path U on M_a at the ground-vertex O of $\bar{O}_1$ are either both orientation-preserving*[11] *or both not.*[12]

PROOF. Let M_b be the cellular decomposition of M dual to M_a. Dual to the vertices and edges of the closed U, in the order as they appear alternately in U, there is on M_b the closed sequence V of incident cells of dimensions n and $n - 1$:

$$V = C_0^n C_0^{n-1} C_1^n C_1^{n-1} \cdots C_t^n C_t^{n-1},$$

where C_0^n is the dual of the vertex O. Let $\bar{M}_b$ denote the cellular decomposition of $\bar{M}$ derived from M_b of M. Then $\bar{M}_b$ is dual to $\bar{M}_a$, and the dual $\bar{V}$ on $\bar{M}_b$ of $\bar{U}$

$$\bar{V} = \bar{C}_0^n \bar{C}_0^{n-1} \bar{C}_1^n \bar{C}_1^{n-1} \cdots \bar{C}_t^n \bar{C}_t^{n-1}$$

covers V in the sense that $\bar{C}_i^n$, $\bar{C}_i^{n-1}$ covers C_i^n, C_i^{n-1}. Since the orientations of $\bar{C}_i$ can be derived from those of C_i, our lemma follows at once.

PROOF OF THEOREM 1. Let the n-complex in Lemma 1 be a closed non-orientable n-manifold M with a definite cellular decomposition. Its only 2-leaved orientable covering manifold M^* is determined by the subgroup G of index 2 of the fundamental group of M, whose elements are the classes of homotopically deformable closed orientation-preserving edge-paths of M at a vertex O^2. A necessary and sufficient condition that a manifold be orientable is obviously that all the closed edge-paths on M at a vertex are orientation-preserving. Suppose that a covering manifold M^{**} of M is orientable. From Lemma 2, its fundamental group is simply isomorphic to a subgroup of G. From Lemma 1, it is a covering manifold of M^*, and hence a covering manifold of M of even number of leaves.

From Lemma 2 and the proof of our theorem, we have the following immediate consequences:

COROLLARY 1. *Any covering manifold of a closed orientable manifold is orientable.*

COROLLARY 2. *Let M be a closed nonorientable manifold and G the group of all classes of homotopically deformable closed orientation-preserving paths on M at a point O. A covering manifold of M, determined by a subgroup of H of the fundamental group of M at O, is orientable when and only when H is a subgroup of G.*

3. PROOF OF THEOREM 2. SUFFICIENCY. Suppose that M is a closed n-manifold, and that $\bar{M}$ a 2-leaved, and therefore regular, orientable covering manifold of M. The covering motion (Deckbewegung) f on $\bar{M}$ is a simplicial

[10] ST, p. 189, p. 272.

[11] ST, p. 191. Notice that an orientation-preserving or orientation-reversing edge-path may have double points.

[12] This lemma is tacitly used in ST, p. 272, in discussion of special M and $\bar{M}$.

topological self-mapping of $\bar{M}$, which is involutory and without fixed point. It remains to show, as follows, that f is orientation-preserving or orientation-reversing according as M is orientable or nonorientable.

Let M_a, m_b, $\bar{M}_a$, $\bar{M}_b$ have the same meaning as in Lemma 2 and its proof, but let M_a be simplicial. Suppose that $\bar{M}$ is determined by the subgroup H of the fundamental group F of M_a with reference to a vertex O of M_a as the initial point of closed edge-paths. Denote by $\bar{O}_1$ and $\bar{O}_2$ the two covering vertices of O. Take an arbitrary closed edge-path U on M_a at O. Denote its covering path on $\bar{M}_a$ at $\bar{O}_1$ by $\bar{U}$. $\bar{U}$ begins at $\bar{O}_1$ and ends at $\bar{O}_2$.

Now let the n-cell on M_b dual to O be C^n, and the n-cells on $\bar{M}_b$ dual to $\bar{O}_i$, $i = 1, 2$, be $\bar{C}_i^n$. The continuation of orientation along $\bar{U}$ can be derived from that along U.[13] When M is orientable (nonorientable), U is orientation-preserving (orientation-reserving). Hence the orientations of $\bar{C}_1^n$ derived from the same orientation (opposite orientations) of C^n are coherent on $\bar{M}_b$. Since f maps $\bar{C}_1^n$ and $\bar{C}_2^n$ onto one another, and their orientations derived from the same orientation of C^n, f is orientation-preserving (orientation-reversing) on $\bar{M}$.

Necessity. Suppose that $\bar{M}$ is a closed orientable n-manifold on which there is a simplicial topological involutory self-mapping f without fixed point. Through identification of pairs of corresponding points on $\bar{M}$ under f, there results a space M. Since the mapping of $\bar{M}$ on M, defined by the identification, is continuous, M is connected. Since f is simplicial and topological in the small, M is a complex and a closed n-manifold respectively. Then obviously $\bar{M}$ fulfills the condition of being a 2-leaved covering manifold of M.

Finally, by virtue of the result in the proof of sufficiency, M is orientable or nonorientable according as f is orientation-preserving or orientation-reversing.

From the fact that the Euler-Poincaré characteristics of M is half that of $\bar{M}$, and from the Poincaré duality theorem for orientable manifold, we have for $n = 2m$ and for $n = 2$ and orientable M the following respectively:

Corollary 3. *If on a closed orientable $(2m)$-manifold there exists a simplicial topological involutory self-mapping without fixed point, the m^{th} Betti number of the manifold must be even.*

Corollary 4. *On a closed orientable 2-manifold of even genus there is no simplicial topological orientation-preserving self-mapping, which is involutory and without fixed point.*[14]

The National University of Peking
and Academia Sinica

[13] ST, p. 271.

[14] Cf. A Komatu, *Über die dreidimensionalen nichtorientierbaren Mannigfaltigkeiten*, Satz 2, Proc. Phys.-Math. Soc. Japan, Vol. 18 (1936), pp. 135–141.

ANNALS OF MATHEMATICS
Vol. 44, No. 2, April, 1943

ON THE NON-EXISTENCE OF REGULAR STATIONARY SOLUTIONS OF RELATIVISTIC FIELD EQUATIONS

BY A. EINSTEIN AND W. PAULI

(Received January 4, 1943)

It is shown that the field equations of the relativistic gravitational theory and of its five-dimensional generalization do not admit any non-singular stationary solution which represents a field of non-vanishing total mass or charge.

Introduction

Some time ago one of us proved[1] that there exist no solutions of the gravitational equations $R_{ik} = 0$ satisfying the following conditions:
1) The field is stationary (i.e. the g_{ik} are independent of x^4).
2) It is free from singularities.
3) It is imbedded in a Euclidean space (of the Minkowski type), and for large values of r (r being the distance from the origin of the spatial coordinate system) g_{44} has the asymptotic form

$$g_{44} = -1 + \frac{\mu}{r},$$

where $\mu \neq 0$.

The third condition implies that the total gravitational mass of the field is different from zero.

The following considerations led us to reanalyze this proof, to reduce it to its necessary elements, and to generalize it to cases of higher dimensions.

When one tries to find a unified theory of the gravitational and electromagnetic fields, he cannot help feeling that there is some truth in *Kaluza's* five-dimensional theory. Yet its foundation is unsatisfactory in so far as, with respect to the group of admissible coordinate transformations, the fifth, space-like, coordinate is treated quite differently from the others. Consequently the components of the electromagnetic field transform independently from those of the gravitational field, and the two fields are only apparently unified. The question arises whether one could base the theory on the full group of five-dimensional point transformations without sacrificing its main achievements.

This might seem impossible, for according to all our experience the physical continuum has 3 + 1 but not 4 + 1 dimensions, since its objects appear to have three, but not four spatial dimensions. One could, however, imagine that this difficulty might be overcome as follows. Assume that in such a theory the fields corresponding to non-singular solutions are not point-like but linearly extended in a four-dimensional space. The geometrical configuration of several coexisting fields of this character would, then, more or less resemble the configuration of the objects of a three-dimensional space.

[1] A. Einstein, Revista (Universidad Nacional de Tucuman) A, 2, 11, 1941.

We therefore have to investigate the question whether, in a five-dimensional metric continuum (of signature 1) the equations

$$R_{ik} = 0$$

admit of non-singular stationary solutions with a field g_{ik} asymptotically given by

$$\begin{matrix} A & 0 & 0 & 0 & 0 \\ 0 & A & 0 & 0 & 0 \\ 0 & 0 & A & 0 & 0 \\ 0 & 0 & 0 & B & D \\ 0 & 0 & 0 & D & C \end{matrix}$$

where at least one of the quantities, A, B, C, D has the form

$$\pm 1 + \frac{\text{const}}{r}$$

with a non-vanishing constant. This is the asymptotic form of a field representing a particle whose electric and ponderable masses do not both vanish.

By discussing the field equations, V. Bargmann[2] has shown that *spherically symmetric* solutions of this character do not exist. In what follows we shall prove the non-existence of regular solutions of the required form irrespective of any assumptions about the symmetry of the field (in regions of finite field strength).

The proof makes it clear why one always encounters singularities if one attempts to represent material particles by solutions of field equations which are based on Riemann's tensor.

1. According to Palatini the variation of the contracted Riemannian curvature tensor[3]

$$R_{IK} = \Gamma^S_{IS,K} - \Gamma^S_{IK,S} + \Gamma^R_{SI}\Gamma^S_{RK} - \Gamma^R_{IK}\Gamma^S_{RS} \tag{1}$$

can be written as

$$\delta R_{IK} = (\delta\Gamma^S_{IS})_{;K} - (\delta\Gamma^S_{IK})_{;S}\,, \tag{2}$$

or

$$\delta R_{IK} = U^S_{IK;S} \tag{3}$$

with

$$U^S_{IK} = -\delta\Gamma^S_{IK} + \tfrac{1}{2}(\delta\Gamma^A_{IA}\delta^S_K + \delta\Gamma^A_{KA}\delta^S_I). \tag{4}$$

This leads to

---

[2] Private communication.

[3] Capital indices assume all values 1, $\cdots$, n, n being the dimension of the space which we consider. Ordinary differentiation is denoted by a comma, covariant differentiation by a semicolon.

(5) $$\sqrt{|g|}\, g^{IK}\delta R_{IK} = \sqrt{|g|}\, g^{IK} U^S_{IK;S},$$

where $|g|$ is the absolute value of the determinant of the covariant metric tensor. Since the covariant derivatives of the metric tensor and of its density vanish, we can put the right side into the form

$$(\sqrt{|g|}\, g^{IK} U^S_{IK})_{;S}$$

and, finally, replace covariant by ordinary differentiation, the expression within the brackets being a contravariant vector density. Hence we get

(6) $$\sqrt{|g|}\, g^{IK}\delta R_{IK} = -\mathfrak{A}^S{}_{,S}$$

with

(7) $$\mathfrak{A}^S = \sqrt{|g|}\,(g^{IK}\delta\Gamma^S_{IK} - g^{AS}\delta\Gamma^B_{AB}).$$

If both the original and the varied fields satisfy the gravitational equations $R_{IK} = 0$, the δR_{IK} in (6) vanish, and we have

(6a) $$\mathfrak{A}^S{}_{,S} = 0.$$

This is always the case when the variation of the field is produced by an infinitesimal change of the coordinate system. In performing such a variation, we must not compare the values of Γ or R at the same world point, but we have to *displace* the world point so as just to compensate the variation of its coordinates due to the change of the coordinate system. Only then can variation and differentiation be interchanged so that we have, for example,

$$(\delta\Gamma^S_{IK})_{,L} = \delta(\Gamma^S_{IK,L}).$$

For a variation produced by the coordinate transformation

$$x'^I = x^I + \xi^I(x)$$

we then have, retaining only terms of the first order in the ξ^I,

(8) $$\delta\Gamma^L_{IK} = -\xi^L{}_{,IK} + \xi^L{}_{,R}\Gamma^R_{IK} - \xi^R{}_{,I}\Gamma^L_{RK} - \xi^R{}_{,K}\Gamma^L_{RI} - \xi^R\Gamma^L_{IK,R}.$$

When (8) is inserted in (7), (6a) must hold for arbitrary functions ξ^I in consequence of the field equations $R_{IK} = 0$.

2. We now decompose the n-dimensional continuum of the x^I into the ordinary three-dimensional space of the x^i (lower case italic indices = 1, 2, 3) and the subspace of the remaining coordinates x^ρ (Greek indices 4, $\cdots$, n). We expressly assume the g_{IK} to be independent of the coordinates x^ρ:

(9) $$\frac{\partial g_{IK}}{\partial x^\rho} = 0.$$

In the physical interpretation of this formalism, x^4 is the time coordinate and x^5 the fifth coordinate introduced in *Kaluza's* theory (where the metric is space-

like with respect to this fifth coordinate). For the following mathematical discussions, however, it is more convenient to restrict neither the dimension of the continuum nor the signature of its metric except for the assumption that the metric of the three-dimensional space be positive definite.

We are going to discuss the implications of the relations (6a) for those variations (8) which do not affect the condition (9) (cylindricity condition). Only the terms with $S = i$ have then to be retained in (6a). Using Gauss' theorem, we infer from $\mathfrak{A}^i{}_{,i} = 0$ that

$$\oint \mathfrak{A}^i n_i \, df = 0 \tag{10}$$

if the integral is extended over a closed two-dimensional surface which does not enclose any singularities of the field. n_i are the covariant components of a unit vector normal to this surface. In case singularities of the metric tensor do exist, we consider two different closed surfaces F_1 and F_2 as inner and outer boundaries of a three-dimensional region free from singularities. Gauss' theorem then leads to

$$\oint_{F_1} \mathfrak{A}^i n_i \, df = \oint_{F_2} \mathfrak{A}^i n_i \, df. \tag{10a}$$

There are two different types of infinitesimal coordinate transformations which leave the stationary character of the field [cf. (9)] invariant. The *first type* is characterized by functions ξ^I which are *independent* of the x^ρ but may depend in an arbitrary way on the x^i. In this case we may choose the ξ^I, together with their first and second derivatives, equal to zero on the inner surface F_1 so that the $\mathfrak{A}^i$ too disappear on F_1. Hence for this type of variations the stronger relation (10) holds even when the surface encloses singularities of the field, provided the surface itself is free from singularities.

The *second type* of transformations not affecting the stationary character of the field leads to an integral theorem which singles out the *regular* solutions of the field equations $R_{IK} = 0$. These transformations are defined by

$$\xi^i = 0, \qquad \xi^\rho = c^\rho_\sigma x^\sigma \tag{11}$$

with constant coefficients c^ρ_σ. From (8) and (9) it then follows that

$$\left.\begin{aligned} &\delta\Gamma^l_{ik} = 0, \qquad \delta\Gamma^B_{iB} = \delta\Gamma^\rho_{i\rho} = 0, \qquad \delta\Gamma^B_{\rho B} = \Gamma^B_{\rho B} = 0 \\ &\delta\Gamma^i_{\rho\sigma} = -(c^\tau_\rho \Gamma^i_{\tau\sigma} + c^\tau_\sigma \Gamma^i_{\rho\tau}), \qquad\qquad \delta\Gamma^l_{i\rho} = -c^\tau_\rho \Gamma^l_{i\tau} \end{aligned}\right\} . \tag{12}$$

Inserting these expressions in (7), we get

$$\mathfrak{A}^i = -2c^\tau_\rho \sqrt{|g|}\, g^{\rho A} \Gamma^i_{A\tau} .$$

Since the coefficients c^τ_ρ may be chosen arbitrarily, (6a) implies, in consequence of the field equations and of the stationary character of the solution,

(13) $$(\sqrt{|g|}\, g^{\rho A}\Gamma^{i}_{A\sigma})_{,i} = 0$$

for all values of ρ and σ. For *non-singular* solutions this leads to

(13a) $$\oint \sqrt{|g|}\, g^{\rho A}\Gamma^{i}_{A\sigma} n_i\, df = 0$$

in analogy with (10), while singular solutions satisfy only a weaker condition of the type (10a).

3. To make use of this theorem, we introduce the assumption that asymptotically the g_{IK}-field in question approaches that of Euclidean space. Hence for large distances from the origin of the coordinate system we may put

(14) $$g_{IK} = \mathring{g}_{IK} + \gamma_{IK}\,.$$

Here the constants $\mathring{g}_{IK}$ can be brought to the form $\delta_{IK}\epsilon_I$ (no summation!) with $\epsilon_i = +1$, $\epsilon_\rho = \pm 1$; γ_{IK} are considered small of the first order. The determinant g is then given by

(15) $$g = (\prod_I \epsilon_I)\cdot(1+\gamma)$$

with

(15a) $$\gamma = \mathring{g}^{IK}\gamma_{IK} = \sum_I \epsilon_I \gamma_{II}\,.$$

Retaining only first order terms, we infer from (1), because of the stationary character of the field, that the equations $R_{IK} = 0$ assume the form

(16) $$\gamma_{IK,ss} - \gamma_{Is,sK} - \gamma_{Ks,sI} + \gamma_{,IK} = 0$$

or

(16a) $$\left.\begin{aligned} \gamma_{ik,ss} - \gamma_{is,sk} - \gamma_{ks,si} + \gamma_{,ik} &= 0 \\ \gamma_{i\rho,ss} - \gamma_{\rho s,si} = 0 \qquad \gamma_{\rho\sigma,ss} &= 0 \end{aligned}\right\}.$$

As is well known, it is always possible to normalize the coordinate system in such a way that

(17) $$\gamma_{Is,s} - \tfrac{1}{2}\gamma_{,I} = 0,$$

or

(17a) $$\gamma_{is,s} - \tfrac{1}{2}\gamma_{,i} = 0 \qquad \gamma_{\rho s,s} = 0$$

then (16) reduces to Laplace's equation

(18) $$\gamma_{IK,ss} = 0.$$

In what follows we consider only those terms which, at infinity, do not decrease more rapidly than r^{-1} (where $r^2 = \sum_i (x^i)^2$). We have, therefore,

(19) $$\gamma_{IK} = m_{IK}/r.$$

Since

$$\gamma = \left(\sum \epsilon_I m_{II}\right) \Big| r = \left(\sum m_{ii} + \sum \epsilon_\rho m_{\rho\rho}\right) \Big| r,$$

we infer from (17a)

$$m_{ik} = \tfrac{1}{2}\delta_{ik}\left(\sum_j m_{jj} + \sum_\rho \epsilon_\rho m_{\rho\rho}\right), \qquad m_{i\rho} = 0,$$

or

$$m_{i\rho} = 0, \qquad m_{ik} = m\delta_{ik}, \qquad m + \sum_\rho \epsilon_\rho m_{\rho\rho} = 0. \tag{20}$$

Hence, neglecting terms of order higher than the first, we have

$$g_{ik} = \delta_{ik}\left(1 + \frac{m}{r}\right), \qquad g_{i\rho} = 0, \qquad g_{\rho\sigma} = \epsilon_\rho \delta_{\rho\sigma} + \frac{m_{\rho\sigma}}{r}, \tag{21}$$

where m satisfies the last equation (20). This result may be used to evaluate the integral (13a) for a sufficiently large sphere. Since on its surface we may put

$$\Gamma^i_{\rho\sigma} = \Gamma_{i,\rho\sigma} = -\tfrac{1}{2}\gamma_{\rho\sigma,i} = -\tfrac{1}{2}m_{\rho\sigma}\left(\frac{1}{r}\right)_{,i} = \tfrac{1}{2}m_{\rho\sigma}\frac{x^i}{r^3} = \tfrac{1}{2}m_{\rho\sigma}\frac{n_i}{r^2}$$

we finally get

$$\oint \sqrt{|g|}\, g^{\rho A}\Gamma^i_{A\sigma} n_i \, df = 2\pi\epsilon_\rho m_{\rho\sigma}. \tag{22}$$

According to this equation, the integral theorems (13a) imply

$$m_{\rho\sigma} = 0 \quad \text{for all} \quad \rho, \sigma, \tag{23}$$

for every solution which is independent of the x^ρ, regular everywhere, and asymptotically approaches the Euclidean metric. Moreover, the relation (20) leads to

$$m = 0. \tag{23a}$$

This shows that for such a non-singular solution of the field equations $R_{IK} = 0$ the deviations of the g_{IK} from the Euclidean (or Minkowskian) $\mathring{g}_{IK}$ must decrease more rapidly than $1/r$ for all values of I and K. Should there exist at all a solution different from the Euclidean metric, it could not describe a particle with *non-vanishing mass or charge*, as stated in the introduction.

Appendix

In the special case of the four-dimensional continuum, R. Serini [Atti Academia dei Lincei (5), 27^1, 235, 1918] has shown that except for the Euclidean metric there exists no regular solution of the required form, under the restrictive assumption, however, that the g_{i4} vanish everywhere. His proof is based on the fact that, under this assumption, we get for the equation (13)

$$(\sqrt{|g|}\, g^{44} g^{ik} g_{44,k})_{,i} = 0.$$

Multiplying this by $-g_{44}$ and integrating over a three-dimensional region we find, because of $g^{44} g_{44} = -1$,

$$\int \sqrt{|g|}\, g^{44} g^{ik} g_{44,i} g_{44,k}\, d^3x + \oint \sqrt{|g|}\, g^{ik} g_{44,k} n_i\, df = 0.$$

According to (13a), the surface integral approaches 0 as the surface moves to infinity, because then $g_{44} \rightarrow -1$, $g_{ik} \rightarrow \delta_{ik}$. Since $g^{44} \neq 0$, $g \neq 0$, and the form $g^{ik} x_i x_k$ is positive definite, the vanishing of the volume integral (extended over the whole space) implies g_{44} = const. For a three-dimensional space, however, the field equations $R_{ik} = 0$ are equivalent to the requirement that the uncontracted Riemannian curvature tensor vanish; hence they imply that the space is Euclidean.

It seems impossible to treat those stationary solutions of the field equations for which the γ_{IK} [i.e. the deviations of the g_{IK} from the constant values $\mathring{g}_{IK}$] decrease more rapidly than $1/r$ by the methods applied in this paper, namely by the use of integral theorems and the linearized field equations. In order to investigate these solutions it will be necessary to discuss in more detail the higher approximations or the exact field equations.

INSTITUTE FOR ADVANCED STUDY

Annals of Mathematics
Vol. 44, No. 2, April 1943

ON THE SUM OF TWO SETS OF INTEGERS

By Emil Artin and Peter Scherk

(Received January 30, 1943)

In his beautiful paper: "A proof of the fundamental theorem on the density of sums of sets of positive integers"[1] Mr. Mann succeeded in proving the (α, β)-hypothesis and a generalization of it that had been conjectured for more than ten years. We found that his method can be simplified considerably and even yields some stronger results.

Let A, B respectively be sets of nonnegative integers a, b. Let $C = A + B$ be the set of all integers of the form $a + b$. Let $A(x)$, $B(x)$, $C(x)$ denote the number of positive integers of the sets $\leqq x$.[2] Mr. Mann proved the following theorem:

If $0 \subset A$ *and* $0 \subset B$, *and if* $C(n) < n$, *then*

$$\frac{C(n)}{n} \geqq \operatorname*{Min}_{\substack{x=1,\cdots,n \\ x \text{ not in } C}} \frac{A(x) + B(x)}{x}. \tag{1}$$

It seems to us remarkable that it is possible to prove a certain identity from which Mr. Mann's inequality can be deduced as an immediate consequence. The theorem in question is the following:

Theorem I. *Let* $n \not\subset C$. *Then*

$$C(n) - C(n - m) = A(m - 1) + B(m - 1) + Z_m \tag{2}$$

for a suitable $m \not\subset C$, $0 < m \leqq n$, *where* Z_m *denotes the number of decompositions of* m *of a certain type.*

Throughout the following proof, small letters always denote numbers between 0 and n, and capital letters stand for sets of such numbers; n is supposed to be not in C. We construct now several sequences of sets denoted by B_ν, B_ν^*, C_ν, and C_ν^*.

Let $C_0 = C_0^* = C$, $B_0 = B_0^* = B$. Let e_1 be the smallest number of B_0^* for which there are two numbers c_1, c_1' not in C_0^* such that

$$c_1 + c_1' - n = a + e_1. \tag{3}$$

With this e_1, we now form C_1 as the set of solutions c_1, c_1' of (3). The corresponding numbers

$$e_1 + n - c_1 = c_1' - a,$$

form the set B_1. Such a number e_1 need not exist; in this case, our construction stops at C_0. C_1 exists if and only if there are elements in C_0^* that have the form $f + f' - n$, where $f, f' \not\subset C_0^*$. The sets C_1^* and B_1^* shall be the union of C_0^* and C_1 respectively of B_0^* and B_1.

---

[1] Annals of Math. 43(1942), pp. 523–527.

[2] Thus, 0 is never counted.

LEMMA 1. $A + B_1^* = C_1^*$.[3]

PROOF: The elements of B_1 can be written in the form $c_1' - a$. So, $A + B_1 \supset C_1$. Since $A + B_0 = C = C_0$, we have $A + B_1^* \supset C_1^*$. So, we have to show that every number $a + b_1 = a + (e_1 + n - c_1)$ belongs to C_1^*, provided it is less than n. In proving this, we may assume that this number does not lie in C_0^*. Calling it c_1', it turns out to be a solution of (3) and, hence, to belong to C_1.

LEMMA 2. $n \not\subset C_1^*$.

It is sufficient to prove $n \not\subset C_1$. Suppose $n = c_1$ in (3) then $c_1' = a + e_1$; from $e_1 \subset B_0^*$ would follow $c_1' \subset C_0^*$.

LEMMA 3. *B_1 and B_0^* are disjoint.*

For if $e_1 + n - c_1 = c_1' - a \subset B_0^*$, c_1' would be of the form $a + b \subset C_0^*$.

Starting from B_1^* and C_1^*, we construct in the same way B_2^*, C_2^*, B_3^*, C_3^*, and so on. This is possible, because B_1^*, C_1^* satisfy still the same conditions as B_0^* and C_0^*, namely Lemma 1 and 2. This process stops, say, at C_h^*. Thus, no number of C_h^* has the form $f + f' - n$, where $f, f' \not\subset C_h^*$. The corresponding numbers e shall be called $e_2, e_3, \cdots, e_h$. Thus, $e_\nu \subset B_{\nu-1}^*$; $n \not\subset C_h^*$, and any two sets B_ν are disjoint.

LEMMA 4. *The numbers e_ν increase monotonically.*[4]

It suffices to show that $e_2 > e_1$. If $e_2 \subset B_0^*$, this follows from the minimum definition of e_1. If $e_2 \subset B_1$, then $e_2 = e_1 + (n - c_1) > e_1$, on account of Lemma 2.

We define $m \leqq n$ as the smallest positive number not in C_h^*.

LEMMA 5. *There are no numbers $c_1 \subset C_1$ with $n - m < c_1 \leqq n - m + e_1$.*

PROOF. Let $c_1 \subset C_1$, $n - m < c_1$. We wish to show $n - m + e_1 < c_1$. This contention is equivalent to $m + c_1 - n > e_1$. Obviously, $0 < m + c_1 - n < m$, hence $m + c_1 - n \subset C_h^*$, say $m + c_1 - n \subset C_\nu$; therefore, because of $C_\nu \subset A + B_\nu : m + c_1 - n = a + b_\nu \geqq b_\nu$.

If $\nu > 0$, then $b_\nu = e_\nu + (n - c_\nu'') > e_\nu \geqq e_1$. If $\nu = 0$, we obtain $m + c_1 - n = a + b$; $c_1' = m$ and c_1 are a solution of (3) with b instead of e_1. Since e_1 was chosen minimal, and since $m \not\subset C_1$, we obtain $b > e_1$.

There are $C_1(n) - C_1(n - m)$ numbers c_1 in the interval $n - m < c_1 < n$. According to Lemma 5, they even satisfy $n - m + e_1 < c_1$ or $e_1 + n - c_1 < m$. These are, according to the definition of B_1, precisely the numbers of B_1 below m. Their number is equal to $B_1(m - 1)$. Thus, we obtain for any $\nu \geqq 1$

$$C_\nu(n) - C_\nu(n - m) = B_\nu(m - 1). \tag{4}$$

LEMMA 6. *All the numbers s in the interval $n - m < s < n$ belong to C_h^*.*

They satisfy, indeed, $0 < s + m - n < m$, so that $s + m - n \subset C_h^*$. If s would not belong to C_h^*, we could construct C_{h+1}^*, for m is also not in C_h^*.

Since $n \not\subset C_h^*$, we obtain from Lemma 6

$$C_h^*(n) - C_h^*(n - m) = m - 1. \tag{5}$$

[3] The sum-sign means: C_1^* is the set of all numbers $a + b_1^*$, where $a \subset A$ and $b_1^* \subset B_1$.

[4] We owe this lemma to a written communication of Professor Alfred Brauer.

There are $m - 1$ decompositions of m into positive summands: $m = x + y$. Among them, there are $A(m - 1)$ with $x \subset A$ and $B_h^*(m - 1)$ with $y \subset B_h^*$. Since $m \not\subset A + B_h^*$, these $A(m - 1) + B_h^*(m - 1)$ decompositions are different from each other. If Z_m denotes the number of decompositions $m = x + y$, where $x \not\subset A$ and $y \not\subset B_h^*$, then

$$m - 1 = A(m - 1) + B_h^*(m - 1) + Z_m .$$

According to Lemma 3, the sets B_ν are disjoint, hence

$$B_h^*(m - 1) = \sum_0^h B_\nu(m - 1)$$

and

$$m - 1 = A(m - 1) + \sum_0^h B_\nu(m - 1) + Z_m . \tag{6}$$

On the other hand, the C_ν are disjoint by construction. So, (5) may be written in the form

$$m - 1 = \sum_0^h (C_\nu(n) - C_\nu(n - m)),$$

or using (4) for $\nu = 1, \cdots, h$

$$m - 1 = C(n) - C(n - m) + \sum_1^h B_\nu(m - 1). \tag{7}$$

By comparing the right sides of (6) and (7), we obtain (2).

As a consequence of Theorem I, we obtain

THEOREM II. *If $C(n) < n$, then*

$$C(n) - C(n - m) \geqq A(m - 1) + B(m - 1) \tag{8}$$

for a suitable $m \not\subset C$ with $0 < m \leqq n$.

Obviously, (2) implies (8) for $n \not\subset C$. If (8) holds for $n - 1$, and if $n \subset C$, then we choose the same m for n as for $n - 1$; the right term of (8) remains unchanged, when we replace $n - 1$ by n, while the left term is not decreased.

If $0 \subset A$, $0 \subset B$, then $m \not\subset B$ and $m \not\subset A$; for if $m \subset A$, we would have $m = 0 + m \subset C$. So, in this case, (8) implies

$$C(n) - C(n - m) \geqq A(m) + B(m). \tag{9}$$

Iterating this formula, we can obtain the following inequality

$$C(n) \geqq \operatorname*{Min}_{\substack{n_0 + \sum m_i = n \\ m_i \text{ not in } C}} (n_0 + \sum (A(m_i) + B(m_i))). \tag{10}$$

Mr. Mann's estimate (1) is an immediate consequence of either (9) or (10).

For

$$n_0 \geqq n_0 \underset{\substack{x=1,\cdots,n \\ x \text{ not in } C}}{\text{Min}} \left(1, \frac{A(x) + B(x)}{x}\right),$$

$$A(m_i) + B(m_i) = m_i \frac{A(m_i) + B(m_i)}{m_i} \geqq m_i \underset{\substack{x=1,\cdots,n \\ x \text{ not in } C}}{\text{Min}} \left(1, \frac{A(x) + B(x)}{x}\right);$$

hence, on account of

$$n_0 + \sum m_i = n,$$

$$C(n) \geqq n \cdot \underset{\substack{x=1,\cdots,n \\ x \text{ not in } C}}{\text{Min}} \left(1, \frac{A(x) + B(x)}{x}\right). \tag{11}$$

Since $C(n) < n$ by assumption,

$$\underset{\substack{x=1,\cdots,n \\ x \text{ not in } C}}{\text{Min}} \left(1, \frac{A(x) + B(x)}{x}\right) = \underset{\substack{x=1,\cdots,n \\ x \text{ not in } C}}{\text{Min}} \frac{A(x) + B(x)}{x},$$

and (1) follows from (11).

Another consequence of (10) is the (α, β)-theorem: Let

$$\alpha > 0, \qquad \beta > 0, \qquad \alpha + \beta < 1; \qquad A(x) \geqq \alpha x, \\ B(x) \geqq \beta x \quad \text{for } x = 1, 2, \cdots, n. \tag{12}$$

Then

$$C(n) \geqq (\alpha + \beta)n.$$

This is clear if $C(n) = n$. But if $C(n) < n$, then (10) yields, on account of $n_0 \geqq (\alpha + \beta)n_0$,

$$C(n) \geqq \underset{\substack{n_0 + \sum m_i = n \\ m_i \text{ not in } C}}{\text{Min}} ((\alpha + \beta)n_0 + \sum (\alpha m_i + \beta m_i))$$

$$= \underset{\substack{n_0 + \sum m_i = n \\ m_i \text{ not in } C}}{\text{Min}} (\alpha + \beta)(n_0 + \sum m_i) = (\alpha + \beta)n.$$

Obviously, (12) can be replaced by the weaker assumption

$$0 < \gamma < 1, \qquad A(x) + B(x) \geqq \gamma x \quad \text{for} \quad x = 1, \cdots, n.$$

Let $0 \subset A$, $0 \not\subset B$. Then C consists of all numbers of the form b and $b + a$, where a $\subset A$, $b \subset B$, and both positive. For such sets, Mr. A. S. Besicovitch has proved[5]: If $A(x) \geqq \alpha(x + 1)$, $B(x) \geqq \beta x$ for $x = 1, \cdots, n$, and if $C(n) < n$, then $C(n) \geqq (\alpha + \beta)n$.

---

[5] *A. S. Besicovitch*: *On the density of the sum of two sets of integers*, Journal London Math. Soc., vol. 10 (1935), pp. 246–248. Mr. Besicovitch's method yields the stronger result: If $A(x) \geqq \alpha(x + 1)$ for $x = 1, \cdots, n$ and if $C(n) < n$, then $C(n) \geqq B(n) + \alpha n$.

Since, in this case, (8) goes over into

$$C(n) - C(n - m) \geqq A(m - 1) + B(m),$$

Mr. Besicovitch's estimate is a consequence of Theorem II. In this case, Theorem II implies inequalities that are analogous to those discussed above.

INDIANA UNIVERSITY

ANNALS OF MATHEMATICS
Vol. 44, No. 2, April, 1943

CONTRIBUTIONS TO THE THEORY OF THE DIRICHLET L-SERIES AND THE EPSTEIN ZETA-FUNCTIONS

BY CARL LUDWIG SIEGEL

Introduction

Let

$$t > 0, \qquad \nu = \left[\sqrt{\frac{t}{2\pi}}\right],$$

$$\vartheta = -\frac{t}{2}\log\pi + \arg\Gamma\left(\frac{1}{4} + \frac{ti}{2}\right) = \frac{t}{2}\log\frac{t}{2\pi e} - \frac{\pi}{8} + O(t^{-1}),$$

$$e^{i\vartheta}\zeta(\tfrac{1}{2} + ti) - 2\sum_{n=1}^{\nu} n^{-\frac{1}{2}}\cos(\vartheta - t\log n) = R.$$

The formula

$$(1) \qquad R = O(t^{-\frac{1}{4}}),$$

due to Hardy and Littlewood[1], is important in the theory of the zeta-function. This formula contains the main term of an asymptotic expansion of $\zeta(\frac{1}{2} + ti)$ for $t \to \infty$, which had been discovered already by Riemann, but was not published[2] before 1932. Riemann's formula is

$$(2) \quad R = (-1)^{\nu-1}\left(\frac{2\pi}{t}\right)^{\frac{1}{4}}\left(C_0 + C_1 t^{-\frac{1}{2}} + \cdots + C_k t^{-\frac{k}{2}} + R_k\right), \qquad R_k = O\left(t^{-\frac{k+1}{2}}\right),$$

where k is an arbitrary integer $\geqq 0$ and $C_0, C_1, \cdots, C_k$ denote certain bounded functions of t, e.g. $C_0 = \cos\left(2\pi u^2 + \frac{3\pi}{8}\right)\Big/\cos 2\pi u$, $u = \sqrt{\frac{t}{2\pi}} - \nu - \frac{1}{2}$. This has been used by Titchmarsh[3], with $k = 1$ and numerical bounds of R_1, for the calculation of the zeros of $\zeta(\sigma + ti)$ in the strip $0 < t < 1468$; he found that all 1041 zeros lie on the critical line $\sigma = \frac{1}{2}$.

Kusmin[4] generalized (1) for the case of an arbitrary L-series, $L(s) = \sum_{n=1}^{\infty}\chi(n)n^{-s}$, where $\chi(n)$ denotes a proper character modulo $m \geqq 1$. In the

[1] G. H. Hardy and J. E. Littlewood, *The zeros of Riemann's zeta-function on the critical line*, Math. Zeitschr. 10, pp. 283–317 (1921).

[2] C. L. Siegel, *Über Riemanns Nachlass zur analytischen Zahlentheorie*, Quell. u. Stud. z. Gesch. d. Math. B2, pp. 45–80 (1932).

[3] E. C. Titchmarsh, *The zeros of the Riemann zeta-function*, Proc. Roy. Soc. London A 151, pp. 234–255 (1935), and 157, pp. 261–263 (1936).

[4] R. O. Kusmin, *Sur les zéros de la fonction $\zeta(s)$ de Riemann*, C. R. Acad. Sci. URSS (N.S.) 2, pp. 398–400 (1934) (Russian. French summary.); R. O. Kusmin, *Zur Theorie der Dirichletschen Reihen* $L(s)$, Bull. Acad. Sci. URSS (7), pp. 1471–1491 (1934) (Russian. German summary).

present paper, I prove the analogue of (2) for $L(s)$; the result (Theorem 6) is of the same form as (2) and contains, of course, (2) as the special case $m = 1$. My proof is somewhat simpler than my former proof of (2). It starts from a representation of $L(s)$ as the sum of two particular integrals (Theorem 4), obtained in a different way by Kusmin; the corresponding theorem for $\zeta(s)$, discovered by Riemann, was published in my edition of Riemann's manuscripts on analytical number-theory. Using a simplification of my former method, remarked by Kusmin, I prove then the following two theorems:

Let $A(t_1, t_2)$ denote the number of different zeros of odd order of $L(s)$ in the interval $t_1 < t < t_2$ on the critical line $s = \frac{1}{2} + ti$. If t_1 is a function of t satisfying the condition $t^{\frac{1}{2}} \log t = o(t - t_1)$, then

$$\liminf_{t \to \infty} A(t_1, t)/(t - t_1) \geqq \gamma, \tag{3}$$

where $\gamma = m/\pi e \varphi(m)$ and $\varphi(m)$ is Euler's function.

Let $B(t_1, t_2, \epsilon)$ denote the number of zeros of $L(s)$ in the rectangle $t_1 < t < t_2$, $\frac{1}{2} - \epsilon < \sigma < \frac{1}{2} + \epsilon$. If $t^{\frac{1}{2}} \log t = o(t - t_1)$ and $\epsilon = o(\log \log t/\log t)$, then

$$\liminf_{t \to \infty} B(t_1, t, \epsilon)/(t - t_1)t^{\epsilon} \geqq \tfrac{1}{4}\gamma. \tag{4}$$

Hardy and Littlewood proved that (3) holds in the case of $\zeta(s)$ with a positive constant γ, provided $t^{\lambda} = O(t - t_1)$ with constant $\lambda > \frac{1}{2}$; however, they did not determine an explicit value of γ. The value $\gamma = m/\pi e \varphi(m)$ is better than the values formerly obtained, in the case $t_1 = 0$, by me for $\zeta(s)$ and by Kusmin for an arbitrary L-function.

The interest of (4) consists in the condition $\epsilon = o(\log \log t/\log t)$ for the breadth of the rectangle. In the well-known theorems of Littlewood[5] and Carlson[6], this breadth is subjected to the conditions $\log \log t/\log t = o(\epsilon)$ and $1 = O(\epsilon)$.

The second part of the paper is concerned with similar problems for certain Epstein zeta-functions, whereas the methods are quite different. Let $Q(x) = Q(x_1, \cdots, x_k)$ be a positive quadratic form of k variables, $\mathfrak{S}$ its matrix, D its determinant and $\zeta(s; \mathfrak{S})$ the Epstein zeta-function defined by the series $\zeta(s; \mathfrak{S}) = \sum Q(n)^{-s} \left(\sigma > \frac{k}{2}\right)$, where $n = (n_1, \cdots, n_k)$ runs over all lattice-points in the k-dimensional space with exception of the origin. The function $\left(s - \frac{k}{2}\right)\zeta(s; \mathfrak{S})$ is regular in the whole plane, and $\eta(s; \mathfrak{S}) = \pi^{-s}\Gamma(s)\zeta(s; \mathfrak{S})$ fulfills the functional equation $\eta(s; \mathfrak{S}) = D^{-\frac{1}{2}}\eta\left(\frac{k}{2} - s; \mathfrak{S}^{-1}\right)$. Obviously $\zeta(s; \mathfrak{C}'\mathfrak{S}\mathfrak{C}) = \zeta(s; \mathfrak{S})$, for any unimodular matrix $\mathfrak{C}$ with k rows; hence $\zeta(s; \mathfrak{S}) = \zeta(s; \mathfrak{S}^{-1})$, if $\mathfrak{S}$ itself

[5] J. E. Littlewood, *On the zeros of the Riemann zeta-function*, Proc. Cambridge Philos. Soc. 22, pp. 295–318 (1925).

[6] F. Carlson, *Über die Nullstellen der Dirichlet'schen Reihen und der Riemann'schen ζ-Funktion*, Ark. Mat. Astr. Fys. no. 20, 28 pp. (1921).

is unimodular. The functional equation shows, in this case, that $\eta(s; \mathfrak{S})$ is real on the line $\sigma = \frac{k}{4}$, which corresponds to the critical line $\sigma = \frac{1}{2}$ for the zeta-function. This holds in particular for $\mathfrak{S} = \mathfrak{E}_k$, the unit matrix of k rows; put $\zeta(s; \mathfrak{E}_k) = \zeta_k(s)$. The formula $\zeta_1(s) = 2\zeta(2s)$ suggests that the distribution of the zeros of $\zeta_k(s)$ might be analogous to that of $\zeta(s)$ itself. Generalizing Hardy's first proof of the existence of an infinite number of zeros of $\zeta(\frac{1}{2} + ti)$, Landau[7] obtained the same result for all $\zeta_k\left(\frac{k}{4} + ti\right)$. He did not notice that, in the special cases $k = 4, 8$, his theorem is an immediate consequence of the formulae

$$(5) \quad \zeta_4(s) = 8(1 - 2^{2-2s})\zeta(s)\zeta(s-1), \quad \zeta_8(s) = 16(1 - 2^{1-s} + 2^{4-2s})\zeta(s)\zeta(s-3)$$

following from Jacobi's theorems on the number of decompositions of an integer into 4 and 8 squares; obviously $\zeta_4(s)$, $\zeta_8(s)$ have on the critical line $\sigma = \frac{k}{4}$ exactly the zeros $s = 1 \pm \frac{l\pi i}{\log 2}$ $(l = 1, 2, \cdots)$, $s = 2 + \frac{i}{\log 2}(2l\pi \pm \operatorname{arc\,tg}\sqrt{15})$ $(l = 0, \pm 1, \cdots)$. For other values of k, however, there is no such obvious reason for the existence of the zeros. My results depend upon two theorems concerning the Epstein zeta-function for arbitrary integral $\mathfrak{S}$:

If $\mathfrak{S}$ and $\mathfrak{S}_1$ belong to the same genus, then

$$(6) \qquad \zeta(s; \mathfrak{S}) - \zeta(s; \mathfrak{S}_1) = t^{\frac{1}{2}} \log t\, O\left(1 + t^{\frac{k}{2} - 2\sigma}\right).$$

Let $\mathfrak{S}_1, \cdots, \mathfrak{S}_h$ be representatives of the different classes of the genus of $\mathfrak{S}$, let $E(\mathfrak{S})$ be the number of units of $\mathfrak{S}$ and

$$Z(s) = \sum_{l=1}^{h} \frac{\zeta(s; \mathfrak{S}_l)}{E(\mathfrak{S}_l)} \Big/ \sum_{l=1}^{h} \frac{1}{E(\mathfrak{S}_l)};$$

define

$$H\left(\frac{a}{b}\right) = D^{-\frac{1}{2}} b^{-\frac{k}{2}} \sum_{n \,(\mathrm{mod}\, b)} e^{\pi i \frac{a}{b} Q(n)},$$

where a, b are coprime positive integers and $abQ(x)$ is an even quadratic form; then

$$(7) \qquad Z(s) = \pi^s \frac{\Gamma\left(\frac{k}{2} - s\right)}{\Gamma\left(\frac{k}{2}\right)} \sum_{a,b} a^{s - \frac{k}{2}} b^{-s} \left\{ e^{\frac{\pi i}{4}(2s-k)} H\left(\frac{a}{b}\right) + e^{\frac{\pi i}{4}(k-2s)} \overline{H\left(\frac{a}{b}\right)} \right\} \qquad \left(1 < \sigma < \frac{k}{2} - 1\right).$$

---

[7] E. Landau, *Über die Hardysche Entdeckung unendlich vieler Nullstellen der Zetafunktion mit reellem Teil* $\frac{1}{2}$, Math. Ann. 76, pp. 212–243 (1915).

The proofs of these two theorems use the theory of modular forms and the analytic theory of quadratic forms. The formulae (6), (7) contain an analogue of (1); it is easy to deduce the following statements:

Let $A(t_1)$ denote the number of different zeros of odd order of $\zeta(s; \mathfrak{S})$ in the interval $0 < t < t_1$ on the line $s = \frac{k}{4} + ti$. If $\mathfrak{S}$ belongs to the genus of $\mathfrak{E}_k$, then

$$A(t) > \frac{t}{\pi}\log 2 + O(1) \qquad (k > 8). \tag{8}$$

Let $B(t_1)$ denote the number of zeros of $\zeta(s; \mathfrak{S})$ in the rectangle $0 < t < t_1$, $2 \leqq \sigma \leqq \frac{k}{2} - 2$. If $\mathfrak{S}$ belongs to the genus of $\mathfrak{E}_k$, then

$$B(t) = \frac{t}{\pi}\log 2 + O(1) \qquad (k \geqq 12). \tag{9}$$

Obviously, (8) and (9) correspond to (3) and (4). The consequences are much more precise; it follows immediately, for $k \geqq 12$, that the zeros of $\zeta(s; \mathfrak{S})$ in the strip $2 \leqq \sigma \leqq \frac{k}{2} - 2$ are simple and lie on $\sigma = \frac{k}{4}$, with at most a finite number of exceptions. In the cases $k = 4, 8$, the zeros of $\zeta(s; \mathfrak{S})$ on $\sigma = \frac{k}{4}$ are completely known, by (5); it is possible to discuss also the remaining cases for $3 < k < 12$, but this requires some numerical computations and we omit it. For the function $\zeta_3(s)$, however, our method does not lead to any result.

Since the number $N(t_1)$ of all zeros in the strip $0 < t < t_1$ satisfies $N(t) = \frac{t}{\pi}\log t + O(t)$, it follows that most of the zeros of $\zeta(s; \mathfrak{S})$ do not lie in the neighborhood of $\sigma = \frac{k}{4}$, if $\mathfrak{S}$ belongs to the genus of $\mathfrak{E}_k$ and $k \geqq 12$.

Part I: Dirichlet L-series

1. Asymptotic expansion

Let σ_1, σ_2 be given real numbers, $\sigma_1 < \sigma_2$, and $s = \sigma + ti$ a complex variable in the half-strip $\sigma_1 \leqq \sigma \leqq \sigma_2$, $t \geqq 2$. If $P, Q \neq 0$ are functions of s and some parameters $p, u, n, \cdots$, then the formula $P = O(Q)$ means that P/Q is bounded in the half-strip, uniformly with respect to the set of values of all parameters $p, u, \cdots$ except n. The symbols $\mathfrak{R}\{c\}$, $\mathfrak{J}\{c\}$ denote real and imaginary part of a complex number c, and $\bar{c}$ is the conjugate complex number. We define

$$\tau = \sqrt{i\left(\frac{1}{2} - s\right)} = \sqrt{t + i\left(\frac{1}{2} - \sigma\right)} = \sqrt{t}\,(1 + O(t^{-1})),$$

$$\sqrt{t} > 0, \qquad \epsilon = e^{\frac{\pi i}{4}} = \frac{1+i}{\sqrt{2}}, \qquad g(z) = e^{\frac{i}{2}z^2} z^{-s}$$

with the principal value of $z^{-s} = e^{-s \log z}$; moreover u is a real variable, p is a positive parameter satisfying the condition $p = \tau + O(1)$ and $c_1, \cdots, c_{21}$ are certain appropriate positive constants.

LEMMA 1:

$$g(p + \epsilon u)/g(\tau) = O(e^{-c_1 u^2}).$$

PROOF: Let $\tau^{-1}(\epsilon u + p - \tau) = v = |v| e^{i\alpha}, -\pi \leqq \alpha < \pi$; then

$$v = t^{-\frac{1}{2}}(\epsilon u + O(1))(1 + O(t^{-1})), \tag{10}$$

whence $\alpha = \arg v = -\frac{\pi}{4} + \frac{\pi}{2} \operatorname{sign} u + O(u^{-1}) + O(t^{-1})$, $\cos \alpha > -\frac{3}{4}$ $(u^2 > c_2;$ $t > c_3)$ and $(\tau v)^2 = iu^2 + (|u| + 1)\, O(1)$; moreover

$$|1 + v| = \frac{p}{|\tau|}\left|1 + \frac{\epsilon}{p} u\right| \geqq \frac{p}{|\tau|\sqrt{2}} > c_4 \qquad (t > c_5). \tag{11}$$

Put

$$\lambda = v + \tfrac{1}{2}v^2 - \log(1 + v) = \int_0^v \frac{w + 2}{w + 1} w\, dw, \qquad e^{-2i\alpha}\lambda = \mu,$$

where the integration is performed over the segment $w = re^{i\alpha}$, $0 \leqq r \leqq |v|$. By (11), $\mu = v^2 O(1)$; on the other hand $\left(r^2 - \frac{9}{4} r + 2\right)(1 + r)^{-2} > c_6$ $(0 \leqq r)$,

$$\begin{aligned}\Re\{\mu\} &= \int_0^{|v|} \frac{r^2 + 3r\cos\alpha + 2}{r^2 + 2r\cos\alpha + 1} r\, dr \\ &\geqq \int_0^{|v|} \left(r^2 - \frac{9}{4} r + 2\right)(1 + r)^{-2} r\, dr \geqq \tfrac{1}{2}c_6 |v|^2 \qquad (u^2 > c_2;\ t > c_3).\end{aligned}$$

Therefore

$$\begin{aligned}\Im\{\tau^2\lambda\} &= \Im\{\tau^2 v^2 |v|^{-2}\mu\} = u^2 |v|^{-2}\Re\{\mu\} + (|u| + 1)|v|^{-2}\mu\, O(1) \\ &\geqq \tfrac{1}{2}c_6 u^2 + O(|u| + 1) > c_1 u^2 \qquad (u^2 > c_7;\ t > c_3)\end{aligned} \tag{12}$$

and

$$\Im\{\tau^2\lambda\} = O(1) \qquad (u^2 \leqq c_7). \tag{13}$$

Since $g(p + \epsilon u) = g(\tau)(1 + v)^{-\frac{1}{2}} e^{i\tau^2\lambda}$, the assertion follows from (11), (12), (13).

Consider now v as an independent complex variable and define

$$h = h(v) = v - \tfrac{1}{2}v^2 - \log(1 + v) \tag{14}$$

with the principal value of $\log(1 + v)$,

$$\psi(v) = (1 + v)^{-\frac{1}{2}} e^{i\tau^2 h}. \tag{15}$$

The function $\psi(v)$ is regular in the circle $|v| < 1$, let

$$\psi(v) = \sum_{n=0}^{\infty} A_n v^n \qquad (|v| < 1)$$

be its power series and

$$S_n(v) = \sum_{k=0}^{n-1} A_k v^k, \qquad R_n(v) = \psi(v) - S_n(v) \qquad (n = 0, 1, \cdots).$$

LEMMA 2: *Let u be real and $v = \tau^{-1}(\epsilon u + p - \tau)$; then*

$$R_n(v) = (|u| + 1)^n O(t^{-\frac{n}{6}}) \qquad (u^2 \leqq c_8 t^{\frac{1}{3}}), \qquad S_n(v) = O(u^n) \qquad (u^2 \geqq c_8 t^{\frac{1}{3}}).$$

PROOF: By (14), the function $v^{-3}h(v)$ is regular in the circle $|v| < 1$ and consequently bounded for $|v| \leqq \frac{1}{2}$, hence $\tau^2 h(v) = O(1)$, $\psi(v) = O(1)$ $(|v| \leqq \frac{1}{2}t^{-\frac{1}{3}})$. Applying the formula

$$|R_n(v)| = \left| \frac{1}{2\pi i} \int_{|z|=\rho} \frac{v^n \psi(z)}{z^n(z - v)}\, dz \right| \leqq \left(\frac{|v|}{\rho}\right)^n \frac{\rho}{\rho - |v|} \max_{|z|=\rho} |\psi(z)| \qquad (|v| < \rho < 1)$$

with $\rho = \frac{1}{2}t^{-\frac{1}{3}}$, $|v| < \frac{1}{2}\rho$, we obtain

$$R_n(v) = v^n O\left(t^{\frac{n}{3}}\right) \qquad (|v| < c_9 t^{-\frac{1}{3}}). \tag{16}$$

Moreover $A_n = v^{-n}(R_{n+1} - R_n) = O\left(|v|\, t^{\frac{n+1}{3}} + t^{\frac{n}{3}}\right)$ $(|v| < c_9 t^{-\frac{1}{3}})$, whence $A_n = O\left(t^{\frac{n}{3}}\right)$,

$$S_n(v) = O\left(1 + |v|^n t^{\frac{n}{3}}\right). \tag{17}$$

By (10), the condition $|v| < c_9 t^{-\frac{1}{3}}$ is satisfied for $v = \tau^{-1}(\epsilon u + p - \tau)$, whenever $u^2 \leqq c_8 t^{\frac{1}{3}}$. The assertion follows from (10), (16), (17).

LEMMA 3: *The coefficients $A_n = A_n(\tau)$ $(n = 0, 1, \cdots)$ of the power series $\psi(v)$ are polynomials in τ^2 of degree $\leqq \frac{n}{3}$ satisfying the recursion formula $nA_n + (n - \frac{1}{2})A_{n-1} + i\tau^2 A_{n-3} = 0$ $(n = 1, 2, \cdots)$ with $A_0 = 1$, $A_{-1} = A_{-2} = 0$.*

PROOF: The function $\psi = \psi(v)$ fulfills the differential equation

$$\frac{d \log \psi}{dv} = -\frac{1}{2}\frac{1}{1 + v} + i\tau^2\left(1 - v - \frac{1}{1 + v}\right),$$

whence

$$(1 + v)\frac{d\psi}{dv} + (\tfrac{1}{2} + i\tau^2 v^2)\psi = 0,$$

$nA_n + (n - 1)A_{n-1} + \frac{1}{2}A_{n-1} + i\tau^2 A_{n-3} = 0$ $(n = 1, 2, \cdots)$, moreover $A_0 = \psi(0) = 1$; q.e.d.

Let $f(u)$ be an integrable function of the real variable u and

$$|f(u)| < c_{10} e^{c_{11} u^2}, \qquad c_{11} < c_1,\ c_{11} < 1. \tag{18}$$

Setting $\epsilon u + p = z$, we define

$$F(s) = \int_{-\infty}^{+\infty} g(z) f(u)\, du, \qquad B_n = \int_{-\infty}^{+\infty} e^{i(z-\tau)^2}(z - \tau)^n f(u)\, du \qquad (n = 0, 1, \cdots).$$

LEMMA 4:

$$F(s)/g(\tau) = \sum_{k=0}^{n-1} A_k B_k \tau^{-k} + O\left(t^{-\frac{n}{6}}\right) \qquad (n = 0, 1, \cdots).$$

PROOF: Introducing again $v = \tau^{-1}(\epsilon u + p - \tau) = \tau^{-1}(z - \tau)$, we have, by (14), (15), $g(z)/g(\tau) = e^{i(z-\tau)^2}\psi(v)$. We use the decomposition

$$F(s)/g(\tau) = \int_{-\infty}^{+\infty} e^{i(z-\tau)^2} S_n(v)f(u)\, du - \int_{|u| \geq a} e^{i(z-\tau)^2} S_n(v)f(u)\, du$$
$$+ \int_{-a}^{a} e^{i(z-\tau)^2} R_n(v)f(u)\, du + \int_{|u| \geq a} g(z)f(u)\, du/g(\tau)$$

with $a^2 = c_8 t^{\frac{1}{3}}$ and obtain, by (18) and Lemmata 1, 2,

$$F(s)/g(\tau) - \sum_{k=0}^{n-1} A_k B_k \tau^{-k} = \int_a^{\infty} e^{-c_{12}u^2} O(u^n)\, du$$
$$+ \int_0^a e^{-c_{12}u^2}(u + 1)^n O\left(t^{-\frac{n}{6}}\right) du + \int_a^{\infty} O(e^{-c_{13}u^2})\, du$$
$$= O(e^{-c_{14}t^{\frac{1}{3}}}) + O\left(t^{-\frac{n}{6}}\right) + O(e^{-c_{15}t^{\frac{1}{3}}}) = O\left(t^{-\frac{n}{6}}\right);$$

q.e.d.

We write $P \approx \sum_{n=0}^{\infty} Q_n \tau^{-n}$ as an abbreviation for the formula $P - \sum_{k=0}^{n-1} Q_k \tau^{-k} = O(\tau^{-n}) (n = 0, 1, \cdots)$. It is easily seen that the relation $P \approx Q$ has the following simple properties: If $P \approx Q$ and $P^* \approx Q^*$, then $P + P^* \approx Q + Q^*$ and $PP^* \approx QQ^*$, where $Q + Q^*$ and QQ^* are sum and product of the formal power series Q, Q^*; if $P \approx Q$, then $P^{-1} \approx Q^{-1}$, whenever the first coefficient Q_0 of Q satisfies the condition $Q_0^{-1} = O(1)$.

THEOREM 1:

$$F(s)/g(\tau) \approx \sum_{k=0}^{\infty} \Gamma_k \tau^{-k},$$

where

$$\Gamma_k = \sum_{l=0}^{3k} \gamma_{kl} \Phi^{(l)}(\tau), \qquad \Phi(y) = \int_{-\infty}^{+\infty} e^{i(z-y)^2} f(u)\, du$$

with numerical coefficients γ_{kl}; in particular $\gamma_{00} = 1$.

PROOF: Since

$$e^{-iy^2}\Phi(\tau + y) = \int_{-\infty}^{+\infty} e^{i(z-\tau)^2 - 2iy(z-\tau)} f(u)\, du,$$

we find

$$(19) \qquad B_k = \int_{-\infty}^{+\infty} e^{i(z-\tau)^2}(z - \tau)^k f(u)\, du = (-2i)^{-k}\{D_y^k e^{-iy^2}\Phi(\tau + y)\}_{y=0},$$

a homogeneous linear function of the derivatives $\Phi^{(l)}(\tau)(l = 0, \cdots, k)$ with numerical coefficients; moreover $z - \tau = \epsilon u + O(1)$ and consequently $\Phi^{(n)}(\tau) = O(1)$ $(n = 0, 1, \cdots)$. On the other hand, by Lemma 3, the expression $A_k\tau^{-k}$ is a polynomial in τ^{-1} of degree $\leqq k$ which does not contain the powers $\tau^{-\nu}$ for $0 \leqq \nu < \frac{k}{3}$. Therefore $\sum_{k=0}^{3n-1} A_k B_k \tau^{-k} = \sum_{k=0}^{3n-1} \Gamma_k^* \tau^{-k}$, where Γ_k^* is a homogeneous linear function of $\Phi^{(l)}(\tau)$ $(l = 0, \cdots, 3k)$ with constant coefficients and independent of n for $k < n$. Writing $\Gamma_k^* = \Gamma_k = \sum_{l=0}^{3k} \gamma_{kl}\Phi^{(l)}(\tau)$ $(k = 0, \cdots, n-1)$ and applying Lemma 4 with $3n$ instead of n, we obtain $F(s)/g(\tau) = \sum_{k=0}^{3n-1} A_k B_k \tau^{-k} + O(\tau^{-n}) = \sum_{k=0}^{n-1} \Gamma_k \tau^{-k} + O(\tau^{-n})$ $(n = 0, 1, \cdots)$, where $\Gamma_0 = \Phi(\tau)$; q.e.d.

2. Properties of the coefficients

In order to get simple recursion formulae for the coefficients γ_{kl}, we introduce the formal power series

$$d_l = d_l(T) = \sum_{k \geqq \frac{l}{3}} \gamma_{kl} T^{-k} \qquad (l = 0, 1, \cdots), \tag{20}$$

T being an indeterminate, and define $d_{-1} = d_{-2} = 0$.

LEMMA 5:

$$\frac{n(n+1)}{2} d_{n+1} + nTd_n + \tfrac{1}{8}d_{n-3} = 0 \qquad (n = 1, 2, \cdots).$$

PROOF: By (19),

$$\sum_{n=0}^{\infty} d_n \Phi^{(n)}(\tau) = \sum_{k=0}^{\infty} \Gamma_k T^{-k} = \sum_{k=0}^{\infty} A_k(T) B_k T^{-k}$$

$$= \sum_{k=0}^{\infty} A_k(T)(-2iT)^{-k}\{D_y^k e^{-iy^2}\Phi(\tau + y)\}_{y=0},$$

whence

$$d_n = \sum_{k=n}^{\infty} A_k(T)(-2iT)^{-k}\binom{k}{n}\{D^{k-n}e^{-iy^2}\}_{y=0}. \tag{21}$$

We write $A_k(T)(-2iT)^{-k}k! = G_k$, $\sum_{k=0}^{\infty} G_k y^{-k} = G$, with another indeterminate y; then the coefficients h_n of the formal Laurent series $L = e^{-iy^2}G = \sum_{n=-\infty}^{\infty} h_n y^{-n}$ are formal power series in T^{-1}, and by (21),

$$d_n = h_n/n! \qquad (n = 0, 1, \cdots). \tag{22}$$

Using Lemma 3, we obtain the recursion formula $TG_{n+1} + \frac{i}{2}\left(n + \frac{1}{2}\right)G_n +$

$\frac{n(n-1)}{8} G_{n-2} = 0$ $(n = 0, 1, \cdots)$; hence G satisfies formally the differential equation

$$T(G-1) - \frac{i}{2}\left(G' - \frac{1}{2y} G\right) + \frac{1}{8}(y^{-1}G)'' = 0.$$

Setting $G = e^{iy^2}L$, we find $(T + \frac{1}{2}y)L + \frac{1}{8}(y^{-1}L)'' = Te^{iy^2}$,

$$Th_n + \tfrac{1}{2}h_{n+1} + \frac{(n-1)(n-2)}{8} h_{n-3} = 0 \qquad (n = 1, 2, \cdots). \tag{23}$$

The assertion follows from (22), (23).

Let

$$\begin{aligned} &\frac{1}{\sqrt{2\pi}}\left(\frac{\tau^2}{2}\right)^{\frac{1}{4}} e^{\frac{\pi}{4}\tau^2} \sqrt{\Gamma\left(\frac{s}{2}\right)\Gamma\left(\frac{1-s}{2}\right)} = \alpha = e^{\omega}, \\ &e^{-\frac{\pi i}{8}}\left(\frac{\tau^2}{2e}\right)^{\frac{i}{2}\tau^2} \sqrt{\Gamma\left(\frac{1-s}{2}\right)\Big/\Gamma\left(\frac{s}{2}\right)} = \beta = e^{-i\theta}. \end{aligned} \tag{24}$$

LEMMA 6:

$$\alpha\beta \approx d_0(\tau), \qquad \alpha^{-2} \approx 1 + \frac{d_1(\tau)}{d_0(\tau)}\tau^{-1}.$$

PROOF: In the special case $f(u) = 1$, we get

$$F(s) = F_0(s) = \int_{-\infty}^{+\infty} g(x)\, du = \pi 2^{\frac{1-s}{2}} e^{\frac{\pi i}{4}s}\Big/\Gamma\left(\frac{s+1}{2}\right),$$

$$\Phi(y) = \Phi_0(y) = \int_{-\infty}^{+\infty} e^{i(z-y)^2}\, du = \sqrt{\pi};$$

consequently, by Theorem 1 and (20),

$$F_0(s)/g(\tau) \approx d_0(\tau)\Phi_0(\tau) = \sqrt{\pi}\, d_0(\tau).$$

For $f(u) = \epsilon u$, we obtain in the same manner

$$F(s) = F_1(s) = \int_{-\infty}^{+\infty} g(z)(z-p)\, du = F_0(s-1) - pF_0(s),$$

$$\Phi(y) = \Phi_1(y) = \int_{-\infty}^{+\infty} e^{i(z-y)^2}(z-p)\, du = (y-p)\int_{-\infty}^{+\infty} e^{i(z-y)^2}\, du = \sqrt{\pi}\,(y-p);$$

$$F_1(s)/g(\tau) \approx d_0(\tau)\Phi_1(\tau) + d_1(\tau)\Phi_1'(\tau) = \sqrt{\pi}\,\{(\tau-p)\, d_0(\tau) + d_1(\tau)\}.$$

Therefore

$$\frac{F_0(s-1)}{F_0(s)} - p \approx \tau - p + \frac{d_1(\tau)}{d_0(\tau)}, \qquad \tau^{-1}\frac{F_0(s-1)}{F_0(s)} \approx 1 + \frac{d_1(\tau)}{d_0(\tau)}\tau^{-1};$$

on the other hand

$$F_0(s-1)/F_0(s) = 2^{\frac{1}{2}} e^{-\frac{\pi i}{4}} \Gamma\left(\frac{s+1}{2}\right) \Big/ \Gamma\left(\frac{s}{2}\right)$$

$$= \pi 2^{\frac{3}{2}} e^{\frac{\pi i}{2}\left(s-\frac{1}{2}\right)} \Big/ (1+e^{\pi i s}) \Gamma\left(\frac{s}{2}\right) \Gamma\left(\frac{1-s}{2}\right)$$

$$= \tau \alpha^{-2}/(1+e^{\pi i s}) = \tau\alpha^{-2}(1+O(e^{-\pi t})).$$

This proves $\alpha^{-2} \approx 1 + \tau^{-1} d_1(\tau)/d_0(\tau)$. Moreover

$$\pi^{-\frac{1}{2}} F_0(s)/g(\tau) = \pi^{\frac{1}{2}} 2^{\frac{1-s}{2}} e^{\frac{\pi i}{4} s} \tau^s e^{-\frac{i}{2}\tau^2} \Big/ \Gamma\left(\frac{s+1}{2}\right) = \alpha\beta(1+e^{\pi i s}),$$

whence $\alpha\beta \approx d_0(\tau)$; q.e.d.

The coefficients a_n, b_n of the power series

$$\frac{1}{\cos x} = \sum_{n=0}^{\infty} \frac{a_n}{(2n)!} x^{2n} \quad \left(|x| < \frac{\pi}{2}\right), \qquad \frac{x}{\sin x} = \sum_{n=0}^{\infty} \frac{b_n}{(2n!)} x^{2n} \quad (|x| < \pi)$$

may be calculated from the recursion formulae

$$\sum_{k=0}^{n} (-1)^k \binom{2n}{2k} a_k = 0, \qquad \sum_{k=0}^{n} (-1)^k \binom{2n+1}{2k} b_k = 0 \qquad (n = 1, 2, \cdots)$$

with $a_0 = b_0 = 1$.

LEMMA 7:

$$\omega \approx \frac{1}{8} \sum_{n=1}^{\infty} \frac{a_n}{n} (2\tau^2)^{-2n}, \qquad \Theta \approx \frac{1}{8} \sum_{n=1}^{\infty} \frac{b_n}{n(2n-1)} (2\tau^2)^{1-2n}.$$

PROOF: Let $\Re\{\xi\} \geqq \frac{1}{4}$, $\xi \neq \frac{1}{4}$; then

$$\frac{d \log \Gamma(\xi)}{d\xi} = \int_0^\infty \left(\frac{e^{-u}}{u} - \frac{e^{-\xi u}}{1-e^{-u}}\right) du = \int_0^\infty (e^{-u} - e^{(\frac{1}{4}-\xi)u}) \frac{du}{u}$$

$$+ \int_0^\infty e^{(\frac{1}{4}-\xi)u} \left(u^{-1} - \frac{e^{-\frac{1}{4}u}}{1-e^{-u}}\right) du$$

$$= \log(\xi - \tfrac{1}{4}) - (4\xi-1)^{-1} + \int_0^\infty e^{(1-4\xi)u} \left(1 + u^{-1} - \frac{4e^{3u}}{e^{4u}-1}\right) du$$

$$\log \Gamma(\xi) = (\xi - \tfrac{1}{2}) \log(\xi - \tfrac{1}{4}) - (\xi - \tfrac{1}{4})$$

$$+ \tfrac{1}{4} \int_0^\infty e^{(1-4\xi)u} \left(\frac{1}{\operatorname{ch} u} + \frac{1}{\operatorname{sh} u} - u^{-1} - 1\right) \frac{du}{u} + c$$

with constant c. On account of the formula $\Gamma(\xi)\Gamma(\xi + \frac{1}{2}) = \sqrt{\pi}\, 2^{1-2\xi}\Gamma(2\xi)$, the passage to the limit $\xi \to \infty$ gives the value $c = \frac{1}{2}\log 2\pi$. Applying Cauchy's theorem, we obtain

$$\log \Gamma\left(\frac{s}{2}\right) = \frac{s-1}{2} \log\left(\frac{s}{2} - \frac{1}{4}\right) - \left(\frac{s}{2} - \frac{1}{4}\right) + \frac{1}{2}\log 2\pi$$

$$+\frac{1}{4}\int_0^\infty e^{i(2s-1)x}\left\{\frac{1}{\cos x}-1+i\left(\frac{1}{\sin x}-x^{-1}\right)\right\}\frac{dx}{x},$$

where the poles $x = \frac{k\pi}{2}$ ($k = 1, 2, \cdots$) of the integrand on the positive real axis are avoided by small half-circles to the left; hence

$$\log \Gamma\left(\frac{s}{2}\right) - \frac{s-1}{2}\log\left(\frac{s}{2}-\frac{1}{4}\right)$$
$$+\left(\frac{s}{2}-\frac{1}{4}\right)-\frac{1}{2}\log 2\pi \approx \frac{1}{8}\sum_{n=1}^{\infty}\left\{\frac{a_n}{n}(2\tau^2)^{-2n}+i\frac{b_n}{n(2n-1)}(2\tau^2)^{1-2n}\right\},$$

and the assertion follows from the definition of ω, Θ in (24).

THEOREM 2:

$$e^{i\Theta}F(s)/g(\tau) \approx \sum_{k=0}^{\infty} G_k\tau^{-k}, \qquad G_k = \sum_{l=0}^{3k} g_{kl}\Phi^{(l)}(\tau), \tag{25}$$

where the coefficients g_{kl} are rational numbers computed from the recursion formulae

$$\delta_{-2} = \delta_{-1} = 0, \qquad \delta_0 = e^{\omega(T)}, \qquad \delta_1 = (\delta_0^{-1} - \delta_0)T, \tag{26}$$

$$\delta_n = -\frac{2T}{n}\delta_{n-1} - \frac{1}{4n(n-1)}\delta_{n-4} \qquad (n = 2, 3, \cdots), \tag{27}$$

$$\omega(T) = \tfrac{1}{8}\sum_{n=1}^{\infty}\frac{a_n}{n}(2T^2)^{-2n}, \qquad a_0 = 1,$$
$$\sum_{k=0}^{n}(-1)^k\binom{2n}{2k}a_k = 0 \qquad (n = 1, 2, \cdots), \tag{28}$$

$$\delta_n = \sum_{k\ge\frac{n}{3}} g_{kn}T^{-k} \qquad (n = 0, 1, \cdots). \tag{29}$$

PROOF: Defining $\Theta(T) = \frac{1}{8}\sum_{n=1}^{\infty}\frac{b_n}{n(2n-1)}(2T^2)^{1-2n}, \delta_n = e^{i\Theta(T)}d_n (n = 0, 1, \cdots)$, we infer from Lemmata 5, 6, 7 that

$$\delta_0 = e^{\omega(T)}, \qquad \delta_0^{-2} = 1 + \frac{\delta_1}{\delta_0}T^{-1},$$

$$\frac{n(n+1)}{2}\delta_{n+1} + nT\delta_n + \tfrac{1}{8}\delta_{n-3} = 0 \qquad (n = 1, 2, \cdots),$$

i.e. (26), (27), (28). Formulae (25), (29) follow from Theorem 1, (20), Lemma 7. By (26), (27), (28), (29), all coefficients g_{kl} are rational numbers.

We find in particular

$$G_0 = \Phi(\tau), \qquad G_1 = -\frac{1}{2^3.3}\Phi^{(3)}(\tau), \qquad G_2 = \frac{1}{2^4}\Phi^{(2)}(\tau) + \frac{1}{2^7.3^2}\Phi^{(6)}(\tau),$$

$$G_3 = -\frac{1}{2^4}\Phi^{(1)}(\tau) - \frac{1}{2^4.3.5}\Phi^{(5)}(\tau) - \frac{1}{2^{10}.3^4}\Phi^{(9)}(\tau),$$

$$G_4 = \frac{1}{2^6}\,\Phi(\tau) + \frac{19}{2^9.3}\,\Phi^{(4)}(\tau) + \frac{11}{2^{11}.3^2.5}\,\Phi^{(8)}(\tau) + \frac{1}{2^{15}.3^5}\,\Phi^{(12)}(\tau).$$

We shall use the symbol $\Omega(\Phi)$ to designate the asymptotic series $\sum_{k=0}^{\infty} G_k \tau^{-k}$.

3. Calculation of some integrals

By the sign $l \swarrow l+1$ we mean that the integration extends over a line $x = \epsilon u + a$, where a is a given real number in the interval $l < a < l+1$, $\epsilon = e^{\frac{\pi i}{4}}$ and u runs through real values from $+\infty$ to $-\infty$. Let q be a positive integer and ξ a complex parameter.

LEMMA 8:

$$\frac{1}{2i}\int_{-1\swarrow 0} e^{q\pi i(x-\xi)^2}\operatorname{ctg}\pi x\,dx = \frac{e^{q\pi i\xi^2}}{1-e^{q\pi i(1-2\xi)}} + \sum_{k=1}^{q}\alpha_k \operatorname{tg}\pi\left(\xi+\frac{k}{q}\right)$$

with constant $\alpha_1, \cdots, \alpha_q$.

PROOF: We denote the integral by J and apply Cauchy's theorem; then

$$\begin{aligned} J - e^{q\pi i\xi^2} &= \frac{1}{2i}\int_{0\swarrow 1} e^{q\pi i(x-\xi)^2}\operatorname{ctg}\pi x\,dx \\ &= \frac{1}{2i}\,e^{q\pi i(1-2\xi)}\int_{-1\swarrow 0} e^{q\pi i(x-\xi)^2+2q\pi ix}\operatorname{ctg}\pi x\,dx \\ &= e^{q\pi i(1-2\xi)}\left\{J + \frac{1}{2i}\int_{-1\swarrow 0} e^{q\pi i(x-\xi)^2}(e^{2q\pi ix}-1)\operatorname{ctg}\pi x\,dx\right\}. \end{aligned}$$

Since

$$\frac{1}{2i}(e^{2q\pi ix}-1)\operatorname{ctg}\pi x = \sum_{k=0}^{q}\beta_k e^{2k\pi ix}, \quad \beta_0 = \beta_q = \tfrac{1}{2}, \quad \beta_k = 1 \qquad (0 < k < q)$$

and

$$\frac{1}{2i}\int_{-1\swarrow 0} e^{q\pi i(x-\xi)^2+2k\pi ix}\,dx = e^{-\pi i\frac{k^2}{q}+2k\pi i\xi}\,b,$$

where

$$(30) \qquad b = \int_{-1\swarrow 0} e^{q\pi i\left(x-\xi+\frac{k}{q}\right)^2}dx = \int_{-1\swarrow 0} e^{q\pi ix^2}\,dx = -\epsilon q^{-\frac{1}{2}}\int_{-\infty}^{+\infty} e^{-\pi x^2}\,dx$$

is independent of ξ, we get

$$(31) \qquad (1-\eta^q)J = e^{q\pi i\xi^2} + \sum_{k=0}^{q}\beta_k^*\,\eta^{q-k}$$

with

$$(32) \qquad \eta = e^{\pi i(1-2\xi)}, \qquad \beta_k^* = (-1)^k e^{-\pi i\frac{k^2}{q}}\beta_k b \qquad (k = 0, \cdots, q).$$

On the other hand,

$$(33)\qquad \sum_{k=0}^{q} \beta_k^* \eta^{q-k}/(1-\eta^q) = \alpha_0 + i\sum_{k=1}^{q} \alpha_k \frac{\eta+\epsilon_k}{\eta-\epsilon_k} = \alpha_0 + \sum_{k=1}^{q} \alpha_k \operatorname{tg} \pi\left(\xi + \frac{k}{q}\right),$$

where $\alpha_0, \alpha_1, \cdots, \alpha_q$ are certain constants and $\epsilon_k = e^{2\pi i \frac{k}{q}}$. Choosing $\eta = 0$ and $\eta = \infty$, we find $2\alpha_0 = \beta_q^* - \beta_0^* = (\beta_q - \beta_0)\, b = 0$, and the assertion follows from (31), (33).

LEMMA 9:

$$\sum_{k=1}^{q} e^{-\pi i \frac{k^2}{q} + \pi i \frac{kn}{q}} = e^{\frac{\pi i}{4}\left(\frac{n^2}{q}-1\right)} q^{\frac{1}{2}} \qquad (n - q \text{ even}).$$

PROOF: Take $\xi = \dfrac{n}{2q}$ and apply (31), (32); then

$$\eta = e^{\pi i\left(1-\frac{n}{q}\right)}, \qquad \eta^q = 1, \qquad b\sum_{k=1}^{q} e^{-\pi i \frac{k^2}{q} + \pi i \frac{kn}{q}} = -\, e^{\pi i \frac{n^2}{4q}},$$

and the assertion follows from (30).

We introduce a proper character $\chi(n)$ modulo $m \geqq 1$ and define

$$W(x) = \frac{\pi}{2m}\sum_{n=1}^{2m} \chi(n) e^{\pi i \frac{n^2}{m}} \operatorname{ctg} \pi \frac{x-n}{2m}, \qquad C = C(\chi) = \sum_{n=1}^{m} \chi(n) e^{-2\pi i \frac{n}{m}}.$$

THEOREM 3:

$$\int_{0 \swarrow 1} e^{\frac{\pi i}{m}x^2 - \frac{2\pi i}{m}\xi x}\, W(x)\, dx = \frac{2\pi i}{1 - e^{-2\pi i \xi}} \sum_{n=1}^{m} \chi(n) e^{-\frac{2\pi i n}{m}\xi} - C e^{-\frac{\pi i}{m}\xi^2}\, \overline{W(\bar{\xi})}.$$

PROOF: Apply Lemma 8 with $q = 4m$, replace x, ξ by $\dfrac{x-n}{2m}, \dfrac{\xi-n}{2m}$, multiply by $2\pi i \chi(n) e^{-\pi i \frac{n^2}{m}}$ and sum over n from 1 to $2m$; then

$$\int_{0 \swarrow 1} e^{\frac{\pi i}{m}(x-\xi)^2}\, W(x)\, dx = \frac{2\pi i}{1 - e^{-4\pi i \xi}} \sum_{n=1}^{2m} \chi(n) e^{-\pi i \frac{n^2}{m} + \frac{\pi i}{m}(\xi - n)^2}$$

$$(34)\qquad + 2\pi i \sum_{n=1}^{2m} \chi(n) e^{-\pi i \frac{n^2}{m}} \sum_{k=1}^{4m} \alpha_k \operatorname{tg} \frac{\pi}{2m}\left(\xi - n + \frac{k}{2}\right)$$

$$= \frac{2\pi i e^{\frac{\pi i}{m}\xi^2}}{1 - e^{-2\pi i \xi}} \sum_{n=1}^{m} \chi(n) e^{-\frac{2\pi i n}{m}\xi} + \sum_{k=1}^{4m} \lambda_k \operatorname{ctg} \frac{\pi}{2m}\left(\xi - \frac{k}{2}\right)$$

with constant $\lambda_1, \cdots, \lambda_{4m}$. Performing the passage to the limit $\xi \to \dfrac{k}{2}$, we find $\lambda_k = 0$ (k odd) and

$$(35)\qquad \frac{2m}{\pi}\lambda_{2l} = -\, e^{\pi i \frac{l^2}{m}} \sum_{n=1}^{m} \chi(n) e^{-2\pi i \frac{ln}{m}} \qquad (l = 1, \cdots, 2m).$$

Since

$$\sum_{n=1}^{m} \chi(n)e^{-2\pi i\frac{ln}{m}} = C\bar{\chi}(l), \tag{36}$$

the assertion follows from (34), (35).

We set

$$\gamma_k = 0 \quad (k - m \text{ odd}), \qquad \gamma_k = \frac{i}{8m} e^{\pi i \frac{k^2}{8m}} \sum_{l=1}^{m} \chi(l) e^{\pi i \frac{l^2}{m} - \pi i \frac{kl}{m}} \qquad (k - m \text{ even}).$$

LEMMA 10:

$$C\bar{\gamma}_k = -e^{-\frac{\pi i}{4}} m^{\frac{1}{2}} \gamma_k \qquad (k = 1, \cdots, 8m).$$

PROOF: The assertion is trivial if $k - m$ is odd. In the other case, by Lemma 9 and (36),

$$\begin{aligned} 8miCe^{\pi i \frac{k^2}{8m}} \bar{\gamma}_k &= \sum_{l,n=1}^{m} \chi(n) e^{-2\pi i \frac{ln}{m} - \pi i \frac{l^2}{m} + \pi i \frac{kl}{m}} \\ &= \sum_{n=1}^{m} \chi(n) e^{\pi i \frac{n^2}{m} - \pi i \frac{kn}{m}} \sum_{l=1}^{m} e^{-\pi i \frac{l^2}{m} + \pi i \frac{kl}{m}} \\ &= -8mie^{-\pi i \frac{k^2}{8m}} \gamma_k e^{\frac{\pi i}{4}\left(\frac{k^2}{m} - 1\right)} m^{\frac{1}{2}}; \end{aligned}$$

q.e.d.

LEMMA 11: *If l is divisible by m, then*

$$\begin{aligned} &\frac{1}{2\pi i} \int_{l \swarrow l+1} e^{\frac{2\pi i}{m}(x-\xi)^2} W(x)\, dx \\ &\quad = \frac{(-1)^l}{1 - (-1)^m e^{-4\pi i\xi}} \sum_{n=1}^{m} \chi(n) e^{-\pi i \frac{n^2}{m} + \frac{2\pi i}{m}(\xi - l - n)^2} + \sum_{k=1}^{8m} \gamma_k \operatorname{ctg} \frac{\pi}{2m}\left(\xi - \frac{k}{4}\right). \end{aligned}$$

PROOF: Apply Lemma 8 with $q = 8m$, replace x, ξ by $\frac{x - n}{2m}, \frac{\xi - n - l}{2m}$, multiply by $\chi(n)e^{-\pi i \frac{n^2}{m}}$ and sum over n from 1 to $2m$; then

$$\begin{aligned} \frac{1}{2\pi i} \int_{0 \swarrow 1} e^{\frac{2\pi i}{m}(x+l-\xi)^2} W(x)\, dx &= \frac{1}{1 - e^{-8\pi i\xi}} \sum_{n=1}^{2m} \chi(n) e^{-\pi i \frac{n^2}{m} + \frac{2\pi i}{m}(\xi - n - l)^2} \\ &\quad + \sum_{n=1}^{2m} \chi(n) e^{-\pi i \frac{n^2}{m}} \sum_{k=1}^{8m} \alpha_k \operatorname{tg} \frac{\pi}{2m}\left(\xi - n - l + \frac{k}{4}\right) \\ &= \frac{e^{\frac{2\pi i}{m}(\xi - l)^2}}{1 - (-1)^m e^{-4\pi i\xi}} \sum_{n=1}^{m} \chi(n) e^{\pi i \frac{n^2}{m} - \frac{4\pi i n}{m}\xi} + \sum_{k=1}^{8m} \gamma_k^* \operatorname{ctg} \frac{\pi}{2m}\left(\xi - \frac{k}{4}\right) \end{aligned} \tag{37}$$

with constant $\gamma_1^*, \cdots, \gamma_{8m}^*$. Performing the passage to the limit $\xi \to \frac{k}{4}$, we

find $\gamma_k^* = 0$ $(k - m$ odd) and

$$(38)\qquad \gamma_k^* = \frac{i}{8m} e^{\frac{2\pi i}{m}\left(\frac{k}{4}-l\right)^2} \sum_{n=1}^{m} \chi(n) e^{\pi i \frac{n^2}{m} - \pi i \frac{nk}{m}} = (-1)^l \gamma_k \qquad (k - m \text{ even}).$$

Since $W(x + l) = (-1)^l W(x)$, the assertion follows from (37), (38).

4. Generalization of Riemann's formulae

It is well-known that C has the absolute value $m^{\frac{1}{2}}$. We define

$$a = \frac{1 - \chi(-1)}{2}, \qquad \rho = i^{-\frac{a}{2}} C^{-\frac{1}{2}} m^{\frac{1}{4}}, \qquad \lambda(s) = \frac{1}{2\pi i} \int_{0 \swarrow 1} e^{\frac{\pi i}{m} x^2} x^{-s} W(x)\, dx,$$

$$\mu(s) = \rho \left(\frac{m}{\pi}\right)^{\frac{s}{2}} \Gamma\left(\frac{s + a}{2}\right) \lambda(s), \qquad L(s) = \sum_{n=1}^{\infty} \chi(n) n^{-s} \qquad (\sigma > 1).$$

THEOREM 4:

$$\rho \left(\frac{m}{\pi}\right)^{\frac{s}{2}} \Gamma\left(\frac{s + a}{2}\right) L(s) = \mu(s) + \overline{\mu(1 - \bar{s})}.$$

PROOF:

$$\int_0^{\epsilon^{-1}\infty} \frac{\xi^{s-1}}{1 - e^{-2\pi i \xi}} \sum_{n=1}^{m} \chi(n) e^{-\frac{2\pi i n}{m}\xi}\, d\xi = \int_0^{\epsilon^{-1}\infty} \xi^{s-1} \sum_{n=1}^{\infty} \chi(n) e^{-\frac{2\pi i n}{m}\xi}\, d\xi$$

$$= \Gamma(s) \sum_{n=1}^{\infty} \chi(n) \left(\frac{2\pi i n}{m}\right)^{-s} = \left(\frac{m}{2\pi}\right)^s e^{-\frac{\pi i}{2}s}\, \Gamma(s) L(s),$$

$$\int_0^{\epsilon^{-1}\infty} \xi^{s-1} \left(\int_{0 \swarrow 1} e^{\frac{\pi i}{m}x^2 - \frac{2\pi i}{m}\xi x} W(x)\, dx\right) d\xi$$

$$= \left(\frac{m}{2\pi}\right)^s e^{-\frac{\pi i}{2}s}\, \Gamma(s) \int_{0 \swarrow 1} e^{\frac{\pi i}{m}x^2} x^{-s} W(x)\, dx;$$

consequently, by Theorem 3 and the formula $W(-x) = -\chi(-1)W(x)$,

$$\left(\frac{m}{2\pi}\right)^s e^{-\frac{\pi i}{2}s}\, \Gamma(s) \{L(s) - \lambda(s)\} = \frac{C}{2\pi i} \int_0^{\epsilon^{-1}\infty} e^{-\frac{\pi i}{m}\xi^2} \xi^{s-1} \overline{W(\bar{\xi})}\, d\xi$$

$$= -\frac{C}{1 + \chi(-1)e^{\pi i s}} \frac{1}{2\pi i} \int_{0 \nwarrow 1} e^{-\frac{\pi i}{m}\xi^2} \xi^{s-1}\, \overline{W(\bar{\xi})}\, d\xi = \frac{C\overline{\lambda(1 - \bar{s})}}{1 + \chi(-1)e^{\pi i s}};$$

moreover

$$2^s e^{\frac{\pi i}{2}s}\, \Gamma\left(\frac{s + a}{2}\right) \Big/ \Gamma(s) = \pi^{-\frac{1}{2}} i^a (1 + \chi(-1)e^{\pi i s}) \Gamma\left(\frac{1 - s + a}{2}\right),$$

and the assertion follows.

Set

$$l = m\left[\sqrt{\frac{t}{2\pi m}}\right], \qquad \eta = y\sqrt{\frac{m}{2\pi}} - l - \frac{1}{2},$$

$$\Phi(y) = \frac{1}{1-(-1)^m e^{-4\pi i\eta}} \sum_{n=1}^{m} \chi(n) e^{-\pi i \frac{n^2}{m} + \frac{2\pi i}{m}\left(\eta+\frac{1}{2}-n\right)^2}$$
$$+ (-1)^l \sum_{k=1}^{8m} \gamma_k \operatorname{ctg} \frac{\pi}{2m}\left(\eta + l + \frac{1}{2} - \frac{k}{4}\right).$$

THEOREM 5:

$$\lambda(s) = \sum_{n=1}^{l} \chi(n) n^{-s} + (-1)^l e^{\frac{i}{2}\tau^2 - i\Theta}\left(\frac{2\pi}{m}\right)^{\frac{s}{2}} \tau^{-s} Q, \qquad Q \approx \Omega(\Phi).$$

PROOF: Putting $\sqrt{\dfrac{m}{2\pi}} = b$, $p = (l + \frac{1}{2})b^{-1}$, $z = \epsilon u + p$, $x = bz$ and $f(u) = -\dfrac{\epsilon b}{2\pi i} W(\frac{1}{2} + \epsilon b u)$, we infer from Cauchy's theorem that

$$\lambda(s) - \sum_{n=1}^{l} \chi(n) n^{-s} = \frac{1}{2\pi i}\int_{l \swarrow l+1} e^{\frac{\pi i}{m}x^2} x^{-s} W(x)\, dx = (-1)^l b^{-s} \int_{-\infty}^{+\infty} g(z) f(u)\, du.$$

By Lemma 11

$$\int_{-\infty}^{+\infty} e^{i(z-y)^2} f(u)\, du = \frac{(-1)^l}{2\pi i} \int_{l \swarrow l+1} e^{\frac{2\pi i}{m}(x-by)^2} W(x)\, dx$$
$$= \frac{1}{1-(-1)^m e^{-4\pi i b y}} \sum_{n=1}^{m} \chi(n) e^{-\pi i \frac{n^2}{m} + \frac{2\pi i}{m}(by-l-n)^2}$$
$$+ (-1)^l \sum_{k=1}^{8m} \gamma_k \operatorname{ctg} \frac{\pi}{2m}\left(by - \frac{k}{4}\right) = \Phi(y),$$

and the assertion follows from Theorem 2.

We define $\chi(n) = e^{i\alpha_n}$, if n and m are coprime, $\rho e^{\frac{\pi i}{8}(2a-1)} = e^{i\delta}$,

$$\Psi(y) = \frac{(-1)^{l-1}}{\sin 2\pi\left(\eta + \frac{m}{4}\right)} \sum_{\substack{n=1\\(n,m)=1}}^{m} \sin\left\{\frac{2\pi}{m}\left(\eta + \frac{m+1}{2} - n\right)^2 - \pi\frac{n^2}{m} + \alpha_n + \delta\right\},$$

$$e^{i\vartheta} = \rho\left(\frac{m}{\pi}\right)^{\frac{s}{2}-\frac{1}{4}} \sqrt{\Gamma\left(\frac{s+a}{2}\right) \Big/ \Gamma\left(\frac{1-s+a}{2}\right)}.$$

THEOREM 6:

$$e^{i\vartheta} L(s) = 2 \sum_{\substack{n=1\\(n,m)=1}}^{l} n^{-\frac{1}{2}} \cos(\vartheta + \alpha_n - \tau^2 \log n) + \left(\frac{2\pi}{m}\right)^{\frac{1}{4}} \tau^{-\frac{1}{2}} R, \qquad R \approx \Omega(\Psi).$$

PROOF: By (24),

$$(39) \quad e^{\frac{i}{2}\tau^2 - i\Theta}\left(\frac{2\pi}{m}\right)^{\frac{s}{2}-\frac{1}{4}} \tau^{\frac{1}{2}-s} = e^{-\frac{\pi i}{8}}\left(\frac{\pi}{m}\right)^{\frac{s}{2}-\frac{1}{4}} \sqrt{\Gamma\left(\frac{1-s}{2}\right) \Big/ \Gamma\left(\frac{s}{2}\right)} = \rho e^{-\frac{\pi i}{8} - i\vartheta} c,$$

where $c = 1$ for $a = 0$ and $c = \operatorname{tg}^{\frac{1}{2}} \dfrac{\pi s}{2}$ for $a = 1$ and consequently $c = i^{\frac{a}{2}} +$

$O(e^{-\pi t})$. The functions

$$\vartheta = \vartheta(s) \text{ and } \nu(s) = \rho\left(\frac{m}{\pi}\right)^{\frac{s}{2}}\Gamma\left(\frac{s+a}{2}\right)e^{-i\vartheta} = \left(\frac{\pi}{m}\right)^{\frac{1}{4}}\sqrt{\Gamma\left(\frac{s+a}{2}\right)\overline{\Gamma\left(\frac{1-\bar{s}+a}{2}\right)}}$$

satisfy $\vartheta(s) = -\overline{\vartheta(1-\bar{s})}$ and $\nu(s) = \overline{\nu(1-\bar{s})}$. It follows from Theorems 4, 5 and (39) that

$$e^{i\vartheta}L(s) = e^{i\vartheta}\lambda(s) + e^{-i\vartheta}\overline{\lambda(1-\bar{s})}$$

$$= 2\sum_{\substack{n=1\\(n,m)=1}}^{l} n^{-\frac{1}{2}}\cos(\vartheta + \alpha_n - \tau^2\log n) + \left(\frac{2\pi}{m}\right)^{\frac{1}{4}}\tau^{-\frac{1}{2}}R,$$

$$R \approx \Omega(\psi), \qquad (-1)^l\psi(y) = \rho e^{\frac{\pi i}{8}(2a-1)}\Phi(y) + \rho^{-1}e^{-\frac{\pi i}{8}(2a-1)}\overline{\Phi(\bar{y})}.$$

By Lemma 10,

$$\rho e^{\frac{\pi i}{8}(2a-1)}\gamma_k + \rho^{-1}e^{-\frac{\pi i}{8}(2a-1)}\overline{\gamma_k} = \rho e^{\frac{\pi i}{8}(2a-1)}(\gamma_k + e^{\frac{\pi i}{4}}Cm^{-\frac{1}{2}}\overline{\gamma_k}) = 0;$$

hence

$$\psi(y) = \frac{(-1)^l}{\sin 2\pi\left(\eta + \frac{m}{4}\right)}\sum_{\substack{n=1\\(n,m)=1}}^{m}$$

$$\sin\left\{\frac{2\pi}{m}(\eta + \tfrac{1}{2} - n)^2 - \pi\frac{n^2}{m} + 2\pi\eta + \pi\frac{m}{2} + \alpha_n + \delta\right\} = \Psi(y);$$

q.e.d.

For our further purposes we need Theorem 4 and the main term in Theorem 5, namely the formula

$$\text{(40)} \qquad \lambda(s) = \sum_{n=1}^{r}\chi(n)n^{-s} + O\left(t^{-\frac{\sigma}{2}}\right), \qquad r = \left[\sqrt{\frac{mt}{2\pi}}\right];$$

obviously

$$\text{(41)} \qquad \log\lambda(s) = \log(L(s) + O(t^{-1})) = \sum_{k,p}\frac{1}{k}\chi(p^k)p^{-ks} + O(t^{-1}) \qquad (\sigma = 3),$$

p running over all prime numbers and k over all positive integers. Moreover, by Lemma 7 and (39),

$$\text{(42)} \qquad \Re\{\vartheta\} = \Re\left\{\frac{\tau^2}{2}\log\frac{m\tau^2}{2\pi e}\right\} + O(1) = \frac{t}{2}\log\frac{mt}{2\pi e} + O(1),$$

$$\text{(43)} \qquad |e^{i\vartheta}| = \left(\frac{mt}{2\pi}\right)^{\frac{\sigma}{2}-\frac{1}{4}}(1 + O(t^{-1})) = O\left(t^{\frac{\sigma}{2}-\frac{1}{4}}\right).$$

5. The zeros of $L(s)$ on and near the critical line

LEMMA 12: *Let $h(s)$ be a regular analytic function on a segment C and δ the variation of* arg $h(s)$ *on C, where the zeros of $h(s)$ on C are avoided by small half-*

circles to the left; let N be the number of different points on C, where the real part of $h(s)$ changes its sign; then

$$N \geqq \frac{\delta}{\pi} - 1. \tag{44}$$

PROOF: Assume first that $h(s) \neq 0$ on C, then $\arg h(s)$ is continuous on C and runs over an interval of length $\geqq |\delta|$. The number of odd multiples of $\frac{\pi}{2}$ in the interior of this interval is $\geqq \frac{|\delta|}{\pi} - 1$; hence $N \geqq \frac{|\delta|}{\pi} - 1 \geqq \frac{\delta}{\pi} - 1$.

Let now $s = s_k (k = 1, \cdots, n)$ be all zeros of $h(s)$ on C, every zero written with its multiplicity; let α be the angle between C and the real axis, $a = e^{i\alpha}$; then $h_1(s) = h(s) \prod_{k=1}^{n} \left(\frac{a}{s - s_k}\right)$ is regular and $\neq 0$ on C, $\Re\{h(s)\} = \Re\{h_1(s)\} \prod_{k=1}^{n} \left(\frac{s - s_k}{a}\right)$. Denoting by δ_1 the variation of $\arg h_1(s)$ on C, we have $\delta_1 = \delta + \pi n$ and therefore, if the real part of $h_1(s)$ changes its sign exactly N_1 times, $N_1 \geqq \frac{\delta_1}{\pi} - 1 = \frac{\delta}{\pi} + n - 1$. On the other hand, $N \geqq N_1 - n$, and (44) follows.

We define $\beta(s) = e^{i\vartheta}\lambda(s) = \rho \left(\frac{m}{\pi}\right)^{\frac{s}{2} - \frac{1}{4}} \sqrt{\Gamma\left(\frac{s + a}{2}\right) \Big/ \Gamma\left(\frac{1 - s + a}{2}\right)}\,\lambda(s)$; by Theorem 4,

$$\tfrac{1}{2} e^{i\vartheta} L(s) = \Re\{\beta(s)\} \qquad (\sigma = \tfrac{1}{2}). \tag{45}$$

LEMMA 13: *Let $\delta(t)$, $\delta_1(t)$ be the variations of* $\arg \beta(s)$, $\arg \lambda(s)$ *on the segment $s = \sigma + ti$, the real part σ running increasingly or decreasingly over a given interval $\sigma_3 \leqq \sigma \leqq \sigma_4$; then $\delta(t) = O(\log t)$, $\delta_1(t) = O(\log t)$.*

PROOF: Consider the function $\gamma(z) = \frac{1}{2}\lambda(z) + \frac{1}{2}\,\overline{\lambda(2ti + \bar{z})}$ in the circle $|z - z_0| \leqq 2r$, $z_0 = 2 + ti$, $r = 2 + |\sigma_3| + |\sigma_4|$. By (40),

$$|\gamma(z)| < \sum_{n < c_{16} t^{\frac{1}{2}}} n^{2r-2} + O(t^{r-1}) = O(t^{r-\frac{1}{2}}),$$

$$\gamma(z_0) = \Re\{\lambda(z_0)\} > 2 - \zeta(2) + O(t^{-\frac{1}{2}}) > c_{17} \qquad (t > c_{18}).$$

On account of Jensen's theorem, the number of zeros of $\gamma(z)$ in the circle $|z - z_0| \leqq r$ is therefore $O(\log t)$. This circle contains the segment $z = \sigma + ti$, $\sigma_3 \leqq \sigma \leqq \sigma_4$, and $\gamma(\sigma + ti) = \Re\{\lambda(\sigma + ti)\}$; consequently the number of zeros of $\Re\{\lambda(\sigma + ti)\}$ on the segment is $O(\log t)$. By Lemma 12, the variation $\delta_1(t)$ of $\arg \lambda(s)$ is $O(\log t)$. By (42), the variation $\delta_2(t)$ of $\arg e^{i\vartheta} = \Re\{\vartheta\}$ is $O(1)$. Since $\delta(t) = \delta_1(t) + \delta_2(t)$, the assertion follows.

We denote by $A(t_1, t_2)$ the number of zeros of odd order of the function $L(\frac{1}{2} + ui)$ in the interval $t_1 < u < t_2$, where $0 < t_1 < t_2$.

LEMMA 14: *Let σ_0 be any number in the interval $\sigma_1 \leqq \sigma < \frac{1}{2}$; then*

$$\pi(\tfrac{1}{2} - \sigma_0) A(t_1, t) > -\int_{t_1}^{t} \log|\beta(\sigma_0 + ui)|\, du + O(\log t).$$

PROOF: Assume that a function $h(s)$ is regular in the rectangle $\sigma_1 \leqq \sigma \leqq \sigma_2$, $t_1 \leqq t \leqq t_2$. We define $\arg h(s)$ on the contour by analytic continuation, beginning at the point $s_1 = \sigma_1 + t_1 i$ and running through the boundary C in positive direction, where zeros of $h(s)$ are avoided by small half-circles to the left. Then

$$\frac{1}{2\pi i}\int_C \log h(s)\, ds = \sum_\rho (s_1 - \rho),$$

where ρ runs over all zeros of $h(s)$ in the interior, whence

$$(46)\qquad \begin{gathered} -\mathfrak{I}\left\{\int_C \log h(s)\, ds\right\} = 2\pi \sum_\rho (\mathfrak{R}\{\rho\} - \sigma_1) \geqq 0, \\ \int_{t_1}^{t_2} \log|h(\sigma_1 + ti)|\, dt + \int_{\sigma_1}^{\sigma_2} \arg h(\sigma + t_2 i)\, d\sigma \\ \geqq \int_{t_1}^{t_2} \log|h(\sigma_2 + ti)|\, dt + \int_{\sigma_1}^{\sigma_2} \arg h(\sigma + t_1 i)\, d\sigma. \end{gathered}$$

Applying this inequality with $h(s) = \beta(s)$, $\sigma_1 = \sigma_0$, $\sigma_2 = \frac{1}{2}$ and with $h(s) = \lambda(s)$, $\sigma_1 = \frac{1}{2}$, $\sigma_2 = 3$, we obtain, by (41) and Lemma 13,

$$(47)\qquad \begin{gathered} \int_{t_1}^{t} \log|\beta(\sigma_0 + ui)|\, du \\ + \int_{\sigma_0}^{\frac{1}{2}} \arg \beta(\sigma + ti)\, d\sigma > \int_{t_1}^{t} \log|\beta(\tfrac{1}{2} + ui)|\, du + O(\log t), \\ \int_{t_1}^{t} \log|\lambda(\tfrac{1}{2} + ui)|\, du > O(\log t). \end{gathered}$$

Let δ be the variation of $\arg \beta(\frac{1}{2} + ui)$ for $t_1 \leqq u \leqq t$; then, by (45) and Lemmata (12), (13),

$$(48)\qquad \arg \beta(\sigma + ti) = \delta + O(\log t) < \pi A(t_1, t) + O(\log t) \quad (\sigma_0 \leqq \sigma \leqq \tfrac{1}{2});$$

moreover $|\beta(\frac{1}{2} + ui)| = |\lambda(\frac{1}{2} + ui)|$. This proves

$$\int_{t_1}^{t} \log|\beta(\sigma_0 + ui)|\, du + \pi(\tfrac{1}{2} - \sigma_0)\, A(t_1, t) > O(\log t);$$

q.e.d.

LEMMA 15: *If* $\sigma_1 \leqq \sigma < \frac{1}{2}$ *and* $t^{\frac{1}{2}} \log t = o(t - t_1)$, *then*

$$(39)\qquad \int_{t_1}^{t} \log|\beta(\sigma + ui)|\, du < \frac{t - t_1}{2}\left(\log \frac{\varphi(m)}{m(1 - 2\sigma)} + o(1)\right).$$

PROOF: Let $t - t_1 = \Delta$, $r = r(t) = \left[\sqrt{\frac{mt}{2\pi}}\right]$, $\sum_{n=1}^{r} \chi(n) n^{-s} = \lambda_0(s)$, then

$$(50)\qquad \int_{t_1}^{t} \log|\beta(\sigma + ui)|\, du \leqq \Delta \log\left(\Delta^{-1} \int_{t_1}^{t} |\beta(\sigma + ui)|\, du\right),$$

and by (40), (43),

$$(51) \qquad \int_{t_1}^{t} |\beta(\sigma + ui)| \, du < \int_{t_1}^{t} \left(\frac{mu}{2\pi}\right)^{\frac{\sigma}{2}-\frac{1}{4}} |\lambda_0(\sigma + ui)| \, du + \Delta O(t^{-\frac{1}{4}}).$$

Suppose first $t < 2t_1$ and set

$$\int_{t_1}^{t} \left(\frac{mu}{2\pi}\right)^{\frac{\sigma}{2}-\frac{1}{4}} |\lambda_0(\sigma + ui)| \, du = H.$$

Since

$$\sum_{n=1}^{r(u)} |\chi(n)|^2 n^{-2\sigma} < \frac{\varphi(m)}{m} \int_0^{r(u)} v^{-2\sigma} \, dv + O(1) = \frac{\varphi(m)}{m(1 - 2\sigma)} \left(\frac{mu}{2\pi}\right)^{\frac{1}{2}-\sigma} (1 + o(1)),$$

we obtain

$$\Delta^{-2} H^2 \leqq \Delta^{-1} \int_{t_1}^{t} \left(\frac{mu}{2\pi}\right)^{\sigma-\frac{1}{2}} |\lambda_0(\sigma + ui)|^2 \, du < \frac{\varphi(m)}{m(1 - 2\sigma)} (1 + o(1))$$

$$+ \Delta^{-1} O\left(\sum_{1 \leqq k < n \leqq r} (kn)^{-\sigma} \left| \int_{t'}^{t} u^{\sigma-\frac{1}{2}} \left(\frac{n}{k}\right)^{iu} du \right| \right),$$

where $t' = \max\left(t_1, \dfrac{2\pi n^2}{m}\right)$, $r = r(t)$. Moreover

$$\int_{t'}^{t} u^{\sigma-\frac{1}{2}} \left(\frac{n}{k}\right)^{iu} du = \frac{1}{\log \frac{n}{k}} O(t^{\sigma-\frac{1}{2}})$$

and

$$\sum_{1 \leqq k < n \leqq r} (kn)^{-\sigma}/\log \frac{n}{k} \leqq \frac{1}{\log 2} \sum_{k \leqq \frac{n}{2}} (kn)^{-\sigma} + \sum_{\frac{n}{2} < k < n} (kn)^{-\sigma} \frac{n}{n - k}$$

$$= O(1) \sum_{n=1}^{r} n^{1-2\sigma} + O(1) \sum_{n=1}^{r} n^{1-2\sigma} \log n = O(t^{1-\sigma} \log t).$$

Hence

$$\Delta^{-2} H^2 < \frac{\varphi(m)}{m(1 - 2\sigma)} (1 + o(1)) + \Delta^{-1} O(t^{\frac{1}{2}} \log t) = \frac{\varphi(m)}{m(1 - 2\sigma)} (1 + o(1)),$$

and the assertion follows from (50), (51).

Consider now the remaining case $t \geqq 2t_1$. We set $\log \frac{t}{t_1} \Big/ \log 2 = h$, $[h] + 1 = h_0$ and apply (49) for the h_0 intervals $u_k \leqq u \leqq u_{k+1}$, $u_k = t_1 2^{\frac{kl}{h_0}}$ $(k = 0, \cdots, h_0 - 1)$; the assertion follows by summation over k.

THEOREM 7: *If* $t^{\frac{1}{2}} \log t = o(t - t_1)$, *then* $\liminf_{t \to \infty} A(t_1, t)/(t - t_1) \geqq m/\pi e \varphi(m)$.

PROOF: Apply Lemmata 14, 15 with $\sigma_0 = \sigma = \frac{1}{2}\left(1 - \frac{e\varphi(m)}{m}\right)$; then

$$\frac{\pi e\varphi(m)}{2m} A(t_1, t) > -\frac{t - t_1}{2} (\log e^{-1} + o(1)) = \frac{t - t_1}{2} (1 + o(1));$$

q.e.d.

We denote by $B(t_1, t_2, \epsilon)$ the number of zeros of $L(s)$ in the rectangle $t_1 < t < t_2$, $\frac{1}{2} - \epsilon < \sigma < \frac{1}{2} + \epsilon$ $(\epsilon > 0, 0 < t_1 < t_2)$.

THEOREM 8: *If $t^{\frac{1}{2}} \log t = o(t - t_1)$ and $\epsilon = o$ $(\log \log t/\log t)$, then*

$$\liminf_{t\to\infty} B(t_1, t, \epsilon)/(t - t_1)t^{\epsilon} \geqq m/4\pi e\varphi(m).$$

PROOF: Let $B^*(t_1, t_2, \epsilon)$ be the number of zeros of $L(s)$ with $\sigma \geqq \frac{1}{2} + \epsilon$, $t_1 < t < t_2$, and put $B(t_1, t, \epsilon) = B$, $B^*(t_1, t, \epsilon) = B^*$, $A(t_1, t) = A$. We have

$$B \geqq A \tag{52}$$

and

$$B + 2B^* = \frac{1}{2\pi} \int_{t_1}^{t} \log \frac{mu}{2\pi} du + O(\log t) > \frac{\Delta}{2\pi} \log \frac{mt}{2\pi e} + O(\log t). \tag{53}$$

On the other hand, by (45),

$$|\beta(\tfrac{1}{2} + ti)| \geqq |\Re\{\beta(\tfrac{1}{2} + ti)\}| = \tfrac{1}{2} |L(\tfrac{1}{2} + ti)|,$$

by (46),

$$\int_{t_1}^{t} \log | L(\tfrac{1}{2} + ui) | du > 2\pi\epsilon B^* + O(\log t).$$

Applying (47), (48) and Lemma 15, we obtain

$$\pi(\tfrac{1}{2} - \sigma)A > 2\pi\epsilon B^* - \Delta \log 2 - \frac{\Delta}{2}\left(\log \frac{\varphi(m)}{m(1 - 2\sigma)} + o(1)\right),$$

where $\sigma_1 \leqq \sigma < \frac{1}{2}$ and $\Delta = t - t_1$. Set $4\pi\epsilon B^*\Delta^{-1} = \eta$ and choose

$$\sigma = \tfrac{1}{2} - \frac{2\varphi(m)}{m} e^{1-\eta}:$$

then

$$A > \frac{m}{4\pi\varphi(m)} \Delta e^{\eta-1}(1 + o(1)).$$

By (52), (53),

$$2\pi\Delta^{-1}B > \max\left\{\frac{m}{2\varphi(m)} e^{\eta-1}(1 + o(1)), \log \frac{mt}{2\pi e} - \epsilon^{-1}\eta\right\} + \Delta^{-1}O(\log t).$$

Putting

$$\eta_0 = \epsilon \log \frac{mt}{2\pi e} - \frac{m\epsilon t^{\epsilon}}{2e\varphi(m)},$$

we infer that

$$\log\frac{mt}{2\pi e}-\epsilon^{-1}\eta \geqq \log\frac{mt}{2\pi e}-\epsilon^{-1}\eta_0=\frac{mt^\epsilon}{2e\varphi(m)} \qquad (\eta \leqq \eta_0),$$

$$e^\eta > e^{\eta_0}=\left(\frac{mt}{2\pi e}\right)^\epsilon(1+o(1)) = t^\epsilon(1+o(1)) \qquad (\eta > \eta_0),$$

whence

$$2\pi\Delta^{-1}B > \frac{mt^\epsilon}{2e\varphi(m)}(1+o(1));$$

q.e.d.

Part II: Epstein zeta-functions

6. Modular forms and Dirichlet series

Let $\Phi(z)$ be a modular form of weight g, with the multiplier system $v = v(a, b, c, d)$ for a subgroup Δ of the modular group Γ, of finite index; this means

$$(cz+d)^{-g}\Phi\left(\frac{az+b}{cz+d}\right)=v\Phi(z) \tag{54}$$

for all substitutions in Δ, and

$$(cz+d)^{-g}\Phi\left(\frac{az+b}{cz+d}\right)=\sum_{n=0}^{\infty}a_n e^{\frac{2\pi i n}{q}z} \qquad (\mathfrak{J}\{z\}>0) \tag{55}$$

for all substitutions in Γ, where the coefficients $a_0, a_1, \cdots$ depend upon a, b, c, d and q is a uniquely determined positive integer. The modular form $\Phi(z)$ is called a cusp-form, if $a_0 = 0$ for all modular substitutions.

With any modular form $\Phi(z)$ we may associate a Dirichlet series $\hat{\Phi}(s)$ in the following way: We consider the expansion (55) in the case of the identical substitution,

$$\Phi(z)=\sum_{n=0}^{\infty}a_n e^{\frac{2\pi i n}{q}z}, \tag{56}$$

and define

$$\hat{\Phi}(s)=\sum_{n=1}^{\infty}a_n n^{-s}. \tag{57}$$

Lemma 16: *Let $\Phi(z)$ be a modular cusp-form of weight g, and let all its multipliers have the absolute value* 1; *then $\hat{\Phi}(s)$ is an entire function and* $\hat{\Phi}(s) = t^{\frac{1}{2}}\log t\, O(1 + t^{g-2\sigma})$.

Proof: It is known that $a_k = O\left(k^{\frac{g}{2}}\right)$, hence (57) converges at least in the half-plane $\sigma > \frac{g}{2}+1$. Let $t > 2$, $\alpha = e^{\pi i(1-t^{-1})}$, $\left(\frac{iq}{2\pi\alpha}\right)^s\Gamma(s)\hat{\Phi}(s)=\varphi(s)$; then

$$(58)\qquad \varphi(s) = \int_0^\infty r^{s-1}\Phi(\alpha r)\,dr.$$

If y denotes the imaginary part of z, then $y^{\frac{g}{2}}\,|\Phi(z)|$ is invariant under the substitutions of Δ, by (54), and vanishes in the parabolic frontier-points of a fundamental domain of Δ, by (55), whence

$$|\Phi(z)| < c_{19}\, y^{-\frac{g}{2}}.$$

Now $\mathfrak{J}\{\alpha r\} = r \sin\frac{\pi}{t} > \frac{2r}{t}$, and consequently

$$(59)\qquad |\Phi(\alpha r)| < c_{19}\left(\frac{t}{2r}\right)^{\frac{g}{2}}.$$

On the other hand, by (56) and the corresponding formula (55) for the substitution $z \to -z^{-1}$, we obtain, since $\mathfrak{J}\{-\alpha^{-1}r^{-1}\} = r^{-1}\sin\frac{\pi}{t} > \frac{2}{rt}$,

$$(60)\qquad |\Phi(\alpha r)| < c_{20}e^{-\frac{4\pi r}{qt}} \quad (r \geqq t), \qquad |\Phi(\alpha r)| < c_{21}r^{-g}e^{-\frac{4\pi}{qrt}} \quad (r \leqq t^{-1}).$$

We use (59), for $t^{-1} \leqq r \leqq t$, and (60). It follows that $\varphi(s)$ is regular in the whole s-plane, by (58), and that

$$\varphi(s) = O\left(\int_0^{t^{-1}} r^{\sigma-g-1}e^{-\frac{4\pi}{qrt}}\,dr + t^{\frac{g}{2}}\int_{t^{-1}}^{t} r^{\sigma-\frac{g}{2}-1}\,dr + \int_t^\infty r^{\sigma-1}e^{-\frac{4\pi r}{qt}}\,dr\right)$$

$$= O\left(t^{g-\sigma} + t^{\frac{g}{2}}\left(t^{\sigma-\frac{g}{2}} + t^{\frac{g}{2}-\sigma}\right)\log t + t^{\sigma}\right) = t^{\sigma}\log t\, O(1 + t^{g-2\sigma}).$$

Since $\left(\frac{2\pi\alpha}{iq}\right)^{s}\Big/\Gamma(s) = O(t^{\frac{1}{2}-\sigma})$, the assertion is proved.

7. Application of the analytic theory of quadratic forms

Let $\mathfrak{S}$ be an integral positive symmetric matrix with k rows and denote by $\alpha(n)$ the number of integral solutions $\mathfrak{x}$ of $\mathfrak{x}'\mathfrak{S}\mathfrak{x} = n$, where n is any integer. It is known that $f(z; \mathfrak{S}) = f(z) = \sum_{n=0}^{\infty}\alpha(n)e^{\pi i n z}$ is a modular form of weight $\frac{k}{2}$ whose multipliers are roots of unity depending upon the genus γ of $\mathfrak{S}$; moreover, if also $\mathfrak{S}_1$ belongs to γ, then the difference $f(z; \mathfrak{S}) - f(z; \mathfrak{S}_1)$ is a cusp-form. The corresponding Dirichlet series $\hat{f}(s) = \sum_{n=1}^{\infty}\alpha(n)n^{-s} = \zeta(s; \mathfrak{S})$ are Epstein zeta-functions. On account of Lemma 16, $\zeta(s; \mathfrak{S}) - \zeta(s; \mathfrak{S}_1) = t^{\frac{1}{2}}\log t\, O\left(1 + t^{\frac{k}{2}-2\sigma}\right)$, whenever $\mathfrak{S}$, $\mathfrak{S}_1$ are in the same genus.

Let $\mathfrak{S}_1, \cdots, \mathfrak{S}_h$ denote representatives of all different classes of γ and define

$$F(z;\gamma) = F(z) = \sum_{l=1}^{h}\frac{f(z;\mathfrak{S}_l)}{E(\mathfrak{S}_l)}\Big/\sum_{l=1}^{h}\frac{1}{E(\mathfrak{S}_l)}, \qquad \hat{F}(s) = Z(s;\gamma),$$

where $E(\mathfrak{S})$ denotes the number of units of $\mathfrak{S}$. Since

$$\zeta(s;\mathfrak{S}) - Z(s;\gamma) = \sum_{l=1}^{h} \frac{\zeta(s;\mathfrak{S}) - \zeta(s;\mathfrak{S}_l)}{E(\mathfrak{S}_l)} \Big/ \sum_{l=1}^{h} \frac{1}{E(\mathfrak{S}_l)},$$

we infer

THEOREM 9:

$$\zeta(s;\mathfrak{S}) = Z(s;\gamma) + t^{\frac{1}{2}} \log t\, O\left(1 + t^{\frac{k}{2}-2\sigma}\right).$$

The matrix $\mathfrak{S}$ is called even, if $\alpha(n) = 0$ for all odd integers n. Let D be the determinant of $\mathfrak{S}$ and define, for any pair of coprime integers a, b and $b > 0$,

$$H\left(\frac{a}{b};\gamma\right) = H\left(\frac{a}{b}\right) = D^{-\frac{1}{2}} b^{-\frac{k}{2}} \sum_{\mathfrak{x} \bmod b} e^{\pi i \frac{a}{b} \mathfrak{x}'\mathfrak{S}\mathfrak{x}},$$

if $ab\,\mathfrak{S}$ is even, and $H\left(\frac{a}{b}\right) = 0$ otherwise.

THEOREM 10: *If* $1 < \sigma < \frac{k}{2} - 1$, *then*

$$Z(s;\gamma) = \pi^s \frac{\Gamma\left(\frac{k}{2} - s\right)}{\Gamma\left(\frac{k}{2}\right)} \sum_{a,b} a^{s-\frac{k}{2}} b^{-s} \left\{e^{\frac{\pi i}{4}(2s-k)} H\left(\frac{a}{b}\right) + e^{\frac{\pi i}{4}(k-2s)} H\left(\frac{-a}{b}\right)\right\},$$

where a, b *run over all pairs of positive coprime integers.*

PROOF: Put $e^{\frac{\pi i}{2}s}\pi^{-s}\,\Gamma(s) Z(s;\gamma) = g(s)$; then

$$g(s) = \int_0^\infty z^{s-1}(F(z) - 1)\, dz \qquad \left(\sigma > \frac{k}{2}\right),$$

where the integration is extended over the positive imaginary axis $z = iy$.

It is known that

$$F(z) - 1 = e^{\frac{\pi i k}{4}} \sum_{a,b} H\left(\frac{a}{b}\right)(bz - a)^{-\frac{k}{2}} \qquad (k < 4), \tag{61}$$

where a, b run over all pairs of coprime integers with $b > 0$, and that $\left|H\left(\frac{a}{b}\right)\right| \leqq 1$. Since the sum $\sum_{a,b} (a^2 + b^2)^{-\frac{k}{4}} \max\left(1, y^{-\frac{k}{2}}\right)$ is a dominant series for the expansion (61) with $z = iy$, we obtain

$$\begin{aligned} g(s) = \frac{H(1)e^{\frac{\pi i}{2}s} Y^{s-\frac{k}{2}}}{s - \frac{k}{2}} &+ e^{\frac{\pi i k}{4}} \sum_{\substack{a,b \\ a\neq 0}} H\left(\frac{a}{b}\right) \int_0^{iY} z^{s-1}(bz - a)^{-\frac{k}{2}}\, dz \\ &+ \int_{iY}^\infty z^{s-1}(F(z) - 1)\, dz \qquad \left(\sigma > \frac{k}{2}\right), \end{aligned} \tag{62}$$

for any $Y > 0$. The last term in (62) is an entire function of s. Now

$$\sum_{\substack{a,b\\a\neq 0}} \int_0^Y y^{\sigma-1}(b^2y^2 + a^2)^{-\frac{k}{4}} dy < \sum_{a\neq 0} \int_0^\infty \left(\int_0^Y y^{\sigma-1}(b^2y^2 + a^2)^{-\frac{k}{4}} dy\right) db \tag{63}$$

$$= 2\zeta\left(\frac{k}{2} - 1\right) \frac{Y^{\sigma-1}}{\sigma - 1} \int_0^\infty (b^2 + 1)^{-\frac{k}{4}} db,$$

for $\sigma > 1$; therefore the infinite series in (62) converges uniformly for $1 < \sigma_1 \leqq \sigma \leqq \sigma_2$ and any fixed $Y > 0$. This proves that (62) holds good for $\sigma > 1$. Assume $1 < \sigma < \frac{k}{2} - 1$ and use (63) with $Y^{-1}, \frac{k}{2} - \sigma$ instead of Y, σ; it follows, for $Y = 1$, that

$$\sum_{\substack{a,b\\a\neq 0}} \int_0^\infty y^{\sigma-1}(b^2y^2 + a^2)^{-\frac{k}{4}} dy$$

$$< 2\zeta\left(\frac{k}{2} - 1\right)\left(\frac{1}{\sigma - 1} + \frac{1}{\frac{k}{2} - \sigma - 1}\right) \int_0^\infty (b^2 + 1)^{-\frac{k}{4}} db,$$

and consequently the series in (62) converges also uniformly with respect to Y. Performing the passage to the limit $Y \to \infty$, we obtain

$$g(s) = e^{\frac{\pi i k}{4}} \sum_{a,b} \left\{H\left(\frac{a}{b}\right) \int_0^\infty z^{s-1}(bz - a)^{-\frac{k}{2}} dz\right.$$

$$\left. + H\left(\frac{-a}{b}\right) \int_0^\infty z^{s-1}(bz + a)^{-\frac{k}{2}} dz\right\} \qquad \left(1 < \sigma < \frac{k}{2} - 1\right),$$

where a, b run over all pairs of positive coprime integers. Since

$$e^{\pi i \left(\frac{k}{2} - s\right)} \int_0^\infty z^{s-1}(bz - a)^{-\frac{k}{2}} dz = \int_0^\infty z^{s-1}(bz + a)^{-\frac{k}{2}} dz = a^{s-\frac{k}{2}} b^{-s} \frac{\Gamma(s)\Gamma\left(\frac{k}{2} - s\right)}{\Gamma\left(\frac{k}{2}\right)},$$

the assertion follows.

Henceforth we assume $D = 1$. If k is not divisible by 8, then there exists exactly one genus, the genus γ_k of the unit matrix $\mathfrak{E}_k$. If k is divisible by 8, then there exist exactly two genera, namely γ_k and a genus γ_k^* of even matrices.

THEOREM 11:

$$Z(s; \gamma_k^*) = \frac{(2\pi)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)\zeta\left(\frac{k}{2}\right)} 2^{-s} \zeta(s) \zeta\left(s + 1 - \frac{k}{2}\right).$$

PROOF: It follows from the known values of the Gaussian sums that $H\left(\frac{a}{b}; \gamma_k^*\right) = 1$. By Theorem 10,

$$Z(s;\gamma_k^*) = 2\pi^s \frac{\Gamma\left(\frac{k}{2}-s\right)}{\Gamma\left(\frac{k}{2}\right)} \cos\frac{\pi s}{2}\,\zeta(s)\,\zeta\left(\frac{k}{2}-s\right)\Big/\zeta\left(\frac{k}{2}\right),$$

and the assertion follows from the functional equation of $\zeta(s)$.

THEOREM 12:

$$(64)\qquad Z(s;\gamma_k) = 2\pi^s \frac{\Gamma\left(\frac{k}{2}-s\right)}{\Gamma\left(\frac{k}{2}\right)} \left\{\psi(s) + \psi\left(\frac{k}{2}-s\right)\right\} \qquad \left(1 < \sigma < \frac{k}{2}-1\right),$$

where

$$\psi(s) = 2^{s-\frac{k}{2}}\left\{\cos\frac{\pi}{4}(2s-k) \sum_{\substack{a,b\\ b\equiv 1 \pmod 4}} \chi_b^k(a) a^{s-\frac{k}{2}} b^{-s} \right.$$
$$\left. + \cos\frac{\pi}{4}(2s+k) \sum_{\substack{a,b\\ b\equiv 3 \pmod 4}} \chi_b^k(a) a^{s-\frac{k}{2}} b^{-s}\right\},$$

$\chi_b(a) = \left(\frac{a}{b}\right)$ *denoting the Legendre-Jacobi symbol.*

PROOF: We have $H\left(\frac{a}{b};\gamma_k\right) = 0$, if ab is odd,

$$(65)\qquad H\left(\frac{a}{b}\right) = i^{k\left(\frac{b-1}{2}\right)^2} \chi_b^k(2a) \qquad (a \text{ even}, b \text{ odd}),$$

$$(66)\qquad H\left(\frac{a}{b}\right) = e^{\frac{\pi i k}{4}} H\left(\frac{-b}{a}\right), \quad H\left(\frac{-a}{b}\right) = e^{-\frac{\pi i k}{4}} H\left(\frac{b}{a}\right) \qquad (0 < a \text{ odd}, b \text{ even}).$$

Define

$$q(s) = \sum_{\substack{a,b\\ a \text{ even}}} a^{s-\frac{k}{2}} b^{-s} \left\{ e^{\frac{\pi i}{4}(2s-k)} H\left(\frac{a}{b}\right) + e^{\frac{\pi i}{4}(k-2s)} H\left(\frac{-a}{b}\right)\right\},$$

then, by Theorem 10 and (66),

$$Z(s;\gamma_k) = \pi^s \frac{\Gamma\left(\frac{k}{2}-s\right)}{\Gamma\left(\frac{k}{2}\right)} \left\{q(s) + q\left(\frac{k}{2}-s\right)\right\}.$$

On the other hand, by (65),

$$q(s) = 2^{s-\frac{k}{2}} \sum_{\substack{a,b\\ b \text{ odd}}} a^{s-\frac{k}{2}} b^{-s} i^{k\left(\frac{b-1}{2}\right)^2} \chi_b^k(a) \left(e^{\frac{\pi i}{4}(2s-k)} + \chi_b^k(-1) e^{\frac{\pi i}{4}(k-2s)}\right) = 2\psi(s);$$

q.e.d.

If k is even, then (64) may be expressed in a different way, analogous to Theorem 11. Obviously

$$\psi(s) = 2^{s-\frac{k}{2}} \cos \frac{\pi}{4}(2s-k) \sum_{\substack{a,b \\ b \text{ odd}}} (-1)^{\frac{k(b-1)}{4}} a^{s-\frac{k}{2}} b^{-s} \qquad (k \text{ even}),$$

whence

$$\psi(s) = 2^{s-\frac{k}{2}} \cos \frac{\pi}{4}(2s-k)\, \zeta\left(\frac{k}{2}-s\right) \zeta(s)(1-2^{-s}) \Big/ \zeta\left(\frac{k}{2}\right)\left(1-2^{-\frac{k}{2}}\right)$$

$$(k \equiv 0 \pmod 4)),$$

$$\psi(s) = 2^{s-\frac{k}{2}} \cos \frac{\pi}{4}(2s-k)\, \zeta\left(\frac{k}{2}-s\right) L_{-4}(s) \Big/ L_{-4}\left(\frac{k}{2}\right) \qquad (k \equiv 2 \pmod 4)),$$

with $L_{-4}(s) = \sum_{n=0}^{\infty} (-1)^n (2n+1)^{-s}$. By (64) and the functional equations of $\zeta(s)$, $L_{-4}(s)$, we obtain

$$Z(s;\gamma_k) = \frac{\pi^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)\zeta\left(\frac{k}{2}\right)\left(1-2^{-\frac{k}{2}}\right)}$$

$$\cdot \left\{1 - 2^{-s} + (-1)^{\frac{k}{4}}\left(2^{\frac{k}{2}-2s} - 2^{-s}\right)\right\} \zeta(s)\,\zeta\left(s+1-\frac{k}{2}\right) \qquad (k \equiv 0),$$

$$Z(s;\gamma_k) = \frac{\pi^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right) L_{-4}\left(\frac{k}{2}\right)}$$

$$\cdot \left\{L_{-4}(s)\,\zeta\left(s+1-\frac{k}{2}\right) + (-1)^{\frac{k-2}{4}} 2^{1-\frac{k}{2}} \zeta(s) L_{-4}\left(s+1-\frac{k}{2}\right)\right\} \qquad (k \equiv 2).$$

For odd values of k, the corresponding formula is more complicated. Let d be a discriminant, i.e. an integer such that either $d \equiv 1 \pmod 4$ and d not divisible by any square $\neq 1$, or $d \equiv 8, 12 \pmod 4$ and $\frac{d}{4}$ not divisible by any square $\neq 1$, and define $L_d(s) = \sum_{n=1}^{\infty} \left(\frac{d}{n}\right) n^{-s}$; then it follows from Theorem 12 and the functional equation of $L_d(s)$ that

$$\pi^{-\frac{k}{2}} \Gamma\left(\frac{k}{2}\right) Z(s;\gamma_k)/\zeta(2s)$$

$$= (1-2^{-2s}) \sum_{d \equiv 1 \pmod 4} d^{\frac{1-k}{2}} L_d\left(s+1-\frac{k}{2}\right) \Big/ L_d\left(s+\frac{k}{2}\right)\left(1-\left(\frac{d}{2}\right)2^{-s-\frac{k}{2}}\right)$$

$$+ \cos\frac{\pi k}{4} 2^{\frac{k}{2}-2s} \sum_{d \equiv 1 \pmod 4} d^{\frac{1-k}{2}} L_d\left(s+1-\frac{k}{2}\right)\left(1-\left(\frac{d}{2}\right)2^{s-\frac{k}{2}}\right)\Big/ L_d\left(s+\frac{k}{2}\right)$$

$$+ \cos\frac{\pi k}{4} 2^{\frac{k}{2}} \sum_{d \equiv 0 \pmod 4} d^{\frac{1-k}{2}} L_d\left(s+1-\frac{k}{2}\right)\Big/ L_d\left(s+\frac{k}{2}\right) \qquad \left(\sigma > \frac{k}{2}\right),$$

where the summation extends over all discriminants d. This formula is of some interest, because it connects the L-series with the theory of the Epstein zeta-functions, but we do not need it for our further purpose and omit the detailed proof.

8. The zeros of $\zeta(s; \mathfrak{S})$ in the strip $1 < \sigma < \frac{k}{2} - 1$.

THEOREM 13: *Let $\mathfrak{S}$ be in the genus γ_k^* and $0 < \epsilon < \frac{k}{4} - 1$; then the number of zeros of $\zeta(s; \mathfrak{S})$ in the strip $1 + \epsilon \leqq \sigma \leqq \frac{k}{2} - 1 - \epsilon$ is finite.*

PROOF: By Theorems 9 and 11, we have

$$\zeta(s; \mathfrak{S}) = Z(s; \gamma_k^*)\left\{1 + t \log t\, O\left(t^{\sigma-\frac{k}{2}} + t^{-\sigma}\right)\right\}$$
$$= Z(s; \gamma_k^*)\{1 + \log t\, O(t^{-\epsilon})\} \qquad \left(1 + \epsilon \leqq \sigma \leqq \frac{k}{2} - 1 - \epsilon\right),$$

and $Z(s; \gamma_k^*)$ has exactly $\frac{k}{4} - 2$ zeros in the strip $1 \leqq \sigma \leqq \frac{k}{2} - 1$, namely $s = 3, 5, \cdots, \frac{k}{2} - 3$.

THEOREM 14: *Let $\mathfrak{S}$ be in the genus γ_k and $k \geqq 12$; then all zeros of $\zeta(s; \mathfrak{S})$ in the strip $2 \leqq \sigma \leqq \frac{k}{2} - 2$ are simple and lie on $\sigma = \frac{k}{4}$, with at most a finite number of exceptions, and the interval $0 < u < t$ contains exactly $\frac{t}{\pi} \log 2 + O(1)$ zeros of $\zeta\left(\frac{k}{4} + ui; \mathfrak{S}\right)$.*

PROOF: By Theorems 9, 12,

$$\zeta(s; \mathfrak{S}) = 2\pi^s \frac{\Gamma\left(\frac{k}{2} - s\right)}{\Gamma\left(\frac{k}{2}\right)} \left\{\psi(s) + \psi\left(\frac{k}{2} - s\right)\right\} \tag{67}$$
$$+ t^{\frac{1}{2}} \log t\, O\left(1 + t^{\frac{k}{2} - 2\sigma}\right) \qquad \left(1 < \sigma < \frac{k}{2} - 1\right),$$

$$\zeta(s; \mathfrak{S}) = \pi^s \frac{\Gamma\left(\frac{k}{2} - s\right)}{\Gamma\left(\frac{k}{2}\right)} e^{\frac{\pi i}{8}(k-4s)} \tag{68}$$

$$\left\{\rho(s) + \overline{\rho\left(\frac{k}{2} - \bar{s}\right)} + \log t\, O(t^{-1})\right\} \qquad \left(2 \leqq \sigma \leqq \frac{k}{2} - 2\right),$$

where

$$\rho(s) = e^{\frac{\pi i k}{8}} 2^{s-\frac{k}{2}} \left\{ \sum_{\substack{a,b \\ b \equiv 1 \pmod 4}} \chi_b^k(a) a^{s-\frac{k}{2}} b^{-s} - i^k \sum_{\substack{a,b \\ b \equiv 3 \pmod 4}} \chi_b^k(a) a^{s-\frac{k}{2}} b^{-s} \right\}.$$

Put $e^{-\frac{\pi i k}{8}} 2^{\frac{k}{2}-s} \rho(s) = 1 + R$, then

$$(69) \qquad 1 + |R| \leqq \sum_{\substack{a,b \\ b \text{ odd}}} a^{\sigma-\frac{k}{2}} b^{-\sigma} = \zeta\left(\frac{k}{2} - \sigma\right)\zeta(\sigma)(1 - 2^{-\sigma})/\zeta\left(\frac{k}{2}\right)\left(1 - 2^{-\frac{k}{2}}\right);$$

and consequently

$$(70) \qquad \rho(s) + \overline{\rho\left(\frac{k}{2} - \bar{s}\right)} = \begin{cases} 2^{s-\frac{k}{2}} e^{\frac{\pi i k}{8}} (1 + R_1) & \left(\sigma = \frac{k}{2} - 2\right) \\ 2^{-s} e^{-\frac{\pi i k}{8}} (1 + R_1) & (\sigma = 2) \end{cases}$$

with

$$|R_1| < -1 + \frac{\zeta(2)\zeta\left(\frac{k}{2} - 2\right)\left(1 - 2^{2-\frac{k}{2}}\right)}{\zeta\left(\frac{k}{2}\right)\left(1 - 2^{-\frac{k}{2}}\right)}$$

$$(71) \qquad + 2^{4-\frac{k}{2}}\left(-1 + \frac{\zeta\left(\frac{k}{2} - 2\right)\zeta(2)(1 - 2^{-2})}{\zeta\left(\frac{k}{2}\right)\left(1 - 2^{-\frac{k}{2}}\right)}\right)$$

$$\leqq \frac{\zeta(2)\zeta(4)}{\zeta(6)(1 - 2^{-6})}(1 + 2^{-3}) - (1 + 2^{-2}) = \frac{3}{4} \qquad (k \geqq 12).$$

Let $\Gamma\left(\frac{k}{2}\right)\pi^{-s} e^{\frac{\pi i}{8}(4s-k)} \zeta(s; \mathfrak{S}) \Big/ \Gamma\left(\frac{k}{2} - s\right) = g(s)$; by (68), (70), (71), the function $\log g(s) - \left(s - \frac{k}{2}\right)\log 2$ is regular and bounded on $\sigma = \frac{k}{2} - 2$ for sufficiently large values of t, and the same holds for $\log g(s) + s \log 2$ on $\sigma = 2$. Analogous to Lemma 13, it is easily proved, by (68), that the variation of $\log g(s)$ on the segment $s = \sigma + ti$ is bounded, if σ runs from 2 to $\frac{k}{2} - 2$. Let $B(t_1)$ be the number of zeros of $\zeta(s; \mathfrak{S})$ in the rectangle $0 < t < t_1$, $2 \leqq \sigma \leqq \frac{k}{2} - 2$; then $2\pi i\, B(t_1)$ is the variation of $\log g(s)$ on the contour, whence

$$(72) \qquad B(t) = \frac{t}{\pi}\log 2 + O(1) \qquad (k \geqq 12).$$

On the other hand, by the functional equation of $\zeta(s; \mathfrak{S})$, we have

$$\zeta(s; \mathfrak{S}) = 2\pi^s \frac{\Gamma\left(\frac{k}{2} - s\right)}{\Gamma\left(\frac{k}{2}\right)} e^{\frac{\pi i}{8}(k-4s)} \Re\{h(s)\} \qquad \left(\sigma = \frac{k}{4}\right),$$

where $h(s)$ is regular on $\sigma = \frac{k}{4}$ and, by (67),

$$h(s) = \rho(s) + \log t\, O(t^{1-\frac{k}{4}}) \qquad (k > 4).$$

By (69),

$$\left| e^{-\frac{\pi i k}{8}} 2^{\frac{k}{2}-s} \rho(s) - 1 \right| \leqq -1 + \zeta^2\left(\frac{k}{4}\right) \Big/ \zeta\left(\frac{k}{2}\right)\left(1 + 2^{-\frac{k}{4}}\right)$$
$$< -1 + \zeta^2(2)/\zeta(4)(1 + 2^{-2}) = 1 \qquad (k > 8);$$

hence

$$\log h(s) = ti \log 2 + O(1) \qquad \left(\sigma = \frac{k}{4}, k > 8\right).$$

Let $A(t_1)$ denote the number of different zeros of odd order of $\zeta(s; \mathfrak{S})$ in the interval $0 < t < t_1$ on the line $s = \frac{k}{4} + ti$. By Lemma 12,

$$A(t) > \frac{t}{\pi} \log 2 + O(1) \qquad (k > 8),$$

and the assertion follows from (72).

THE INSTITUTE FOR ADVANCED STUDY

ANNALS OF MATHEMATICS
Vol. 44, No. 2, April, 1943

ON CONTINUOUS PATH-SURFACES OF ZERO AREA

BY TIBOR RADÓ

(Received December 28, 1942)

INTRODUCTION

0.1. If a curve C, in terms of Cartesian coordinates x_1, x_2, x_3, is given by equations of the form

$$(1) \qquad x_i = x_i(t), \qquad i = 1, 2, 3, \qquad t' \leqq t \leqq t'',$$

then the points (x_1, x_2, x_3), obtained by means of these equations, form a certain point-set E. However, the length of the curve cannot be deduced from a mere knowledge of E. For example, we may assign E as a straight segment which is described, for suitable choice of the equations (1), several times in either direction while t varies from t' to t''. Hence it is clear that the length of a *path-curve*, that is a curve determined by equations of the form (1), is determined by the manner in which the point-set E is described, rather than by the point-set E itself. The purpose of the term *path-curve* is precisely to call attention to this fact.

0.2. Similar considerations apply to surfaces. A *continuous path-surface* S is determined, in terms of Cartesian coordinates x_1, x_2, x_3, by equations of the form

$$(2) \qquad S: \quad x_i = x_i(P), \qquad i = 1, 2, 3, \qquad P \epsilon U,$$

where $x_1(P)$, $x_2(P)$, $x_3(P)$ are continuous functions of the point P on the surface U of the unit sphere. The term path-surface has been proposed by C. B. Morrey (see Morrey [1] in the Bibliography at the end of this paper). In the literature on the area of surfaces, the range of definition of the coordinate functions is usually taken as a plane Jordan region, rather than the surface U of the unit sphere. The discussion is quite analogous in either case, but it is more convenient for our present purposes to work on the unit sphere. It will be also convenient to use vector notation. Instead of (2), we shall write

$$(3) \qquad S: \quad \mathfrak{x} = \mathfrak{x}(P), \qquad P \epsilon U,$$

where $\mathfrak{x}$ stands for the vector with components x_1, x_2, x_3.

0.3. Suppose there exists a triangulation T of U, comprised of (curvilinear) triangles $\Delta_1, \Delta_2, \cdots, \Delta_m$, such that each Δ_j is carried, by means of (3), in a biunique way into a plane rectilinear triangle Δ_j' in (x_1, x_2, x_3) space. In this case, we shall call S a *polyhedral path-surface*, and we shall designate by $E(S)$ the area of S in the elementary sense, that is, the sum of the areas of the triangles Δ_j'.

0.4. It will be convenient to consider also another type of polyhedral surfaces, to be termed simply *polyhedra*, in contradistinction to the polyhedral path-surfaces defined in 0.3. Let again T be a triangulation of U, comprised of the

(curvilinear) triangles $\delta_1, \delta_2, \cdots, \delta_n$. With each vertex V of the triangulation T let there be associated a point V' in (x_1, x_2, x_3) space. With each triangle δ_j of T we associate the rectilinear triangle δ_j', in (x_1, x_2, x_3) space, whose vertices are the three points V' which were associated with the three vertices V of δ_j. The triangles δ_j' constitute then a *polyhedron* $\mathfrak{P}$ and the sum of the areas of $\delta_1', \delta_2', \cdots, \delta_n'$ is the area $E(\mathfrak{P})$ of $\mathfrak{P}$. The usefulness of these definitions will become apparent in the sequel. It should be observed that we do not require that the correspondence between the vertices V and V' be bi-unique. As a consequence, some or all of the triangles δ_j' may reduce to straight segments or even to single points. If this happens for a certain triangle δ_j', then its area is taken as equal to zero.

0.5. The definition of the Lebesgue area $A(S)$ of the continuous path-surface (3) may now be stated in the following two forms which are easily shown to be equivalent. Let

$$S_n: \qquad \mathfrak{x} = \mathfrak{x}_n(P), \qquad P \in U,$$

be a sequence of polyhedral path-surfaces such that

$$\mathfrak{x}_n(P) \to \mathfrak{x}(P)$$

uniformly on U. Then

$$A(S) = \text{gr. l. b. } \underline{\lim}\, E(S_n), \tag{4}$$

where the greatest lower bound is taken with respect to all sequences S_n with the properties just stated. The reader interested in learning about recent literature on the Lebesgue area $A(S)$, may consult Radó [2].

0.6. Given a continuous path-surface S as in (3), and a polyhedron $\mathfrak{P}$, let us introduce a quantity $d(S, \mathfrak{P})$ which will in a sense measure the deviation of $\mathfrak{P}$ from S, as follows. According to 0.4, $\mathfrak{P}$ arises from a triangulation T of U, comprised of certain curvilinear triangles $\delta_1, \delta_2, \cdots, \delta_n$. For each vertex V of T, we have a corresponding vertex V' of $\mathfrak{P}$. On the other hand, to each vertex V of T there corresponds a point $\bar{V}$ by means of (3). We define $d(S, \mathfrak{P})$ as the maximum distance between the points V' and $\bar{V}$ for all vertices V of T, *plus* the maximum diameter of the triangles $\delta_1, \delta_2, \cdots, \delta_n$ of T. We have then the formula, easily recognized as equivalent to (4),

$$A(S) = \text{gr. l. b. } \underline{\lim}\, E(\mathfrak{P}_n),$$

where the greatest lower bound is taken with respect to all sequences of polyhedra $\mathfrak{P}_n$ such that $d(S, \mathfrak{P}_n) \to 0$.

0.7. It is obvious that a continuous path-curve has zero length if and only if it reduces to a single point. The analogous question for continuous path-surfaces, to be referred to for brevity as *the characterization problem*, is of considerable complexity, and also of considerable importance as an apparently indispensable step in approaching various unsolved fundamental problems in the theory of the area of surfaces. The characterization problem has been

studied extensively by Geöcze (see Geöcze) [1], [2], [3], [4]). His main contributions are described in sections 3.16, 2.9, 2.14, 1.21, 3.5 of this paper. The recent work of Morrey on generalized conformal maps of surfaces (Morrey [1]) implies several interesting results concerning our topic. For example, the theorems in 2.8 and 2.13 could be also deduced from the general results of Morrey. Since his work depends to a large extent upon conformal mapping and Lebesgue theory, it is interesting to observe that our methods are essentially topological, except for some very elementary details. The principal contributions of the present paper, as far as new results are concerned, are contained in the main theorem in 2.12 and in the fundamental lemma in 2.1.

0.8. As far as the general plan of approach is concerned, the present study is based upon the work of Geöcze (see bibliography). However, our results are more comprehensive, and our statements and methods are far less involved. We shall make extensive use of two fundamental topological conceptions which were not available to Geöcze: first, the conception of *the topological index*, or order, of a point with respect to a plane closed continuous path-curve, and second, the conception of a *cyclic element* of a Peano space. The application of the latter to our problem was suggested by the work of Morrey who was the first to apply the structure theory of Peano spaces (see Kuratowski-Whyburn [1], also for further references) in the theory of the area of surfaces (see Morrey [1]). While the topological index has been a standard tool in the theory of the area for many years, various useful modifications were suggested to the author by the work of Eilenberg on the topology of the plane (Eilenberg [1]).

0.9. For the convenience of the reader, we collect here some of the most frequently used symbols and terms, to avoid excessive cross-references.

SYMBOLS. U, 0.2; $E(S)$, 0.3; $E(\mathfrak{P})$, 0.4; $A(S)$, 0.5, 0.6; $d(S, \mathfrak{P})$, 0.6; Σ, 1.1; ω, 1.20.

TERMS. Continuous path-surface, 0.2; polyhedral path-surface, 0.3; polyhedron, 0.4; non-degenerate vector-function, 1.3; single-valued continuous argument, 1.5; condition (Arg), 1.6; indicator continuum, 1.7; indicator curve, 1.8; associated Peano space, 1.16; dendrite, 1.19.

In the preceding list, each item is followed by the number of the section where that item is explained.

CHAPTER I. PRELIMINARIES

1.1. In the sequel, only metric spaces will be needed. We shall use Σ to designate a *Peano space*, that is a space which is compact, connected and locally connected.

1.2. Let $x_1(P)$, $x_2(P)$, $x_3(P)$ be three continuous real-valued functions, where the point P varies on a Peano space Σ. We shall designate by $\mathfrak{x}(P)$ the continuous vector-function with components $x_1(P)$, $x_2(P)$, $x_3(P)$, and we shall write

$$\mathfrak{x}(P) = (x_1(P), x_2(P), x_3(P)), \qquad P \in \Sigma.$$

1.3. The continuous vector-function $\mathfrak{x}(P)$ is *non-degenerate* on Σ if it is not constant on any continuum containing more than one point.

1.4. We shall associate with the continuous vector-function $\mathfrak{x}(P)$ three continuous complex-valued functions as follows:

$$f_1(P) = x_2(P) + ix_3(P),$$
$$f_2(P) = x_3(P) + ix_1(P),$$
$$f_3(P) = x_1(P) + ix_2(P).$$

1.5. Let $f(P)$ be a continuous complex-valued function on the Peano-space Σ. Let C be a continuum on Σ, such that $f(P) \neq 0$ on C. If there exists on C a single-valued, continuous, real-valued function $\varphi(P)$ such that

$$f(P) = |f(P)| (\cos \varphi(P) + i \sin \varphi(P)), \qquad P \in C,$$

then $\varphi(P)$ will be called a *single-valued continuous argument* for $f(P)$ on C. The following facts are well known.

1) If $\varphi(P)$, $\psi(P)$ are two single-valued continuous arguments for $f(P)$ on C, then on C the difference $\varphi(P) - \psi(P)$ is equal to a constant of the form $2k\pi$, k an integer.

2) If $\varphi(P)$ is a single-valued continuous argument for $f(P)$ on C, and if k is an integer, then $\varphi(P) + 2k\pi$ is also a single-valued continuous argument for $f(P)$ on C.

3) If C is an arc (topological image of the interval $0 \leqq t \leqq 1$), and if $f(P) \neq 0$ on C, then $f(P)$ possesses a single-valued continuous argument on C.

1.6. Let $f(P)$ be a complex-valued continuous function on a Peano space Σ, and let C be a continuum on Σ. We shall say that $f(P)$ satisfies condition (Arg) on C, if for every choice of the complex constant ζ, such that $f(P) \neq \zeta$ on C, the function $f(P) - \zeta$ possesses a single-valued continuous argument on C. Here and in the sequel, the term *complex* is used to mean *complex or real,* as usual.

1.7. Let $\mathfrak{x}(P)$ be a continuous vector-function on a Peano space Σ. A continuum C on Σ will be called an *indicator continuum* for $\mathfrak{x}(P)$, if at least one of the associated functions $f_1(P)$, $f_2(P)$, $f_3(P)$, defined in 1.4, fails to satisfy condition (Arg) on C.

1.8. If an indicator continuum is a simple closed curve (topological image of the perimeter of the unit circle), then C will be called an *indicator curve* for $\mathfrak{x}(P)$.

1.9. The definition of an indicator curve may be re-stated in the following equivalent form. Let C be a simple closed curve on Σ. Let us assign a positive direction on C. The equation

$$\mathfrak{x} = \mathfrak{x}(P), \qquad P \in C$$

defines then a *closed continuous path-curve* C' in (x_1, x_2, x_3) space. Let us introduce the complex variables

$$\zeta_1 = x_2 + ix_3, \qquad \zeta_2 = x_3 + ix_1, \qquad \zeta_3 = x_1 + ix_2.$$

If we project C' orthogonally upon the ζ_1, ζ_2, ζ_3 planes respectively, we obtain three plane closed continuous path-curves C_1', C_2', C_3', which are represented by the equations (cf. 1.4)

$$\begin{aligned} C_1': \quad & \zeta_1 = f_1(P), \qquad P \in C, \\ C_2': \quad & \zeta_2 = f_2(P), \qquad P \in C, \\ C_3': \quad & \zeta_3 = f_3(P), \qquad P \in C. \end{aligned}$$

Let ζ_i^0 be a point of the plane ζ_i. Then the *topological index* (or order) of ζ_i^0 with respect to C_i' is defined as follows. If ζ_i^0 is on C_i', then the index is zero. If ζ_i^0 is not on C_i', then the continuously varying argument of the function $f_i(P) - \zeta_i^0$ changes by an amount $2n\pi$, n an integer, while P describes C in the positive sense. The integer n is then the index of ζ_i^0 with respect to C_i' (if we reverse the positive direction on C, then the index changes merely its sign). The index, with respect to C_i', is thus a function defined for all points of the plane ζ_i. Let us designate it by $\mu_i(\zeta_i)$. This function really depends also upon $\mathfrak{x}(P)$ and C, but it is unnecessary for our purposes to use more complicated notations.

Clearly then, the simple closed curve C is an indicator curve for $\mathfrak{x}(P)$, in the sense of 1.8, if and only if at least one of the index-functions $\mu_1(\zeta_1)$, $\mu_2(\zeta_2)$, $\mu_3(\zeta_3)$ is not identically equal to zero.

1.10. Lemma. *On a Peano space Σ, let there be given a continuous complex-valued function $f(P)$. Suppose there exists a single-valued continuous argument for $f(P)$ on every simple closed curve on which $f(P) \neq 0$. Then there also exists a single-valued continuous argument for $f(P)$ on every domain (connected open set) on which $f(P) \neq 0$.*

Proof. Let D be any domain on which $f(P) \neq 0$. In D we pick a point P_0, and an argument φ_0 of the complex number $f(P_0)$. We shall keep P_0 and φ_0 fixed. Let P_1 be any point of D different from P_0, and let γ be any arc that joins P_0 and P_1 in D. On γ we have a single-valued continuous argument $\varphi(P;\gamma)$ of $f(P)$, such that $\varphi(P_0;\gamma) = \varphi_0$ (cf. 1.5). Clearly, our lemma is proved if we can show that

$$\varphi(P_1;\gamma') = \varphi(P_1;\gamma'') \tag{5}$$

for any two arcs γ'; γ'' which join P_0 and P_1 in D.

To prove (5), we denote by E the set of all common points of γ' and γ''. Then E is a closed subset of γ' which contains (at least) the points P_0 and P_1. We consider the function

$$\psi(P) = \frac{\varphi(P;\gamma'') - \varphi(P;\gamma')}{2\pi}, \qquad P \in E.$$

Then $\psi(P)$ is well-defined and continuous on E, and it assumes only integral values on E. Indeed, for every point $P \in E$, $\varphi(P;\gamma'')$ and $\varphi(P;\gamma')$ are arguments

of the same complex number, namely, of $f(P)$. Hence, $\psi(P)$ is an integer. We have

$$\psi(P_0) = 0.$$

Let us assume now, in contradiction with (5), that

$$\psi(P_1) \neq 0.$$

Let us designate by E_0, E_1 the subsets of E on which $\psi(P) = 0$, $\psi(P) \neq 0$ respectively. From the properties of $\psi(P)$ described above, it follows that E_0, E_1 are two non-vacuous closed sets on γ' without common points. It follows that we have a sub-arc γ_0' of γ', bounded by two (distinct) points Q_0, Q_1, such that

$$(6) \qquad Q_0 \in E_0, \qquad Q_1 \in E_1, \qquad E \cdot \gamma_0' = Q_0 + Q_1 .$$

These same two points Q_0, Q_1 are the endpoints of a sub-arc γ_0'' of γ'', and from (6) it follows that

$$(7) \qquad \gamma_0' \cdot \gamma_0'' = Q_0 + Q_1 .$$

Hence γ_0' and γ_0'' form a simple closed curve C_0 on which $f(P) \neq 0$, since $C_0 \subset D$. By assumption, we have a single-valued continuous argument $\varphi(P)$ for $f(P)$ on C_0. We have then (cf. 1.5)

$$\varphi(P) - \varphi(P; \gamma') = \text{constant on } \gamma_0',$$

$$\varphi(P) - \varphi(P; \gamma'') = \text{constant on } \gamma_0''.$$

Hence

$$\varphi(Q_0) - \varphi(Q_0 ; \gamma') = \varphi(Q_1) - \varphi(Q_1 ; \gamma'),$$

$$\varphi(Q_0) - \varphi(Q_0 ; \gamma'') = \varphi(Q_1) - \varphi(Q_1 ; \gamma'').$$

Subtraction yields

$$\psi(Q_0) = \psi(Q_1).$$

This is a contradiction, since $Q_0 \in E_0$, $Q_1 \in E_1$.

1.11. Lemma. *On a Peano space Σ, let there be given a continuous complex-valued function $f(P)$. Suppose there exists a single-valued continuous argument for $f(P)$ on every simple closed curve on which $f(P) \neq 0$. Then there also exists a single-valued continuous argument for $f(P)$ on every continuum on which $f(P) \neq 0$.*

Proof. Let Γ be any continuum on which $f(P) \neq 0$. Since Σ is locally connected and since $f(P)$ is continuous, there exists, for every point $Q \in \Gamma$, a connected open set $D(Q)$, containing Q, on which $f(P) \neq 0$. Let D denote the sum of all the sets $D(Q)$ which are associated in this manner with the points Q of Γ. Then D is a domain, and $f(P) \neq 0$ on D. By 1.10, we have a single-valued continuous argument for $f(P)$ on D and hence also on Γ, since $\Gamma \subset D$.

1.12. Lemma. *Let $\mathfrak{x}(P)$ be a continuous vector function on a Peano space Σ. If there exists an indicator continuum for $\mathfrak{x}(P)$ on Σ, then there also exists an indicator curve for $\mathfrak{x}(P)$ on Σ.*

This lemma is a direct consequence of 1.11.

1.13. Let $\mathfrak{x}(P)$ be a continuous vector-function on a Peano space Σ. Then Σ is the sum of continua, without common points, on each of which $\mathfrak{x}(P)$ is constant, and each of which is maximal with respect to this property. Let K designate the class of these continua. Then, in the terminology of R. L. Moore, K is an upper semi-continuous collection of continua, filling up Σ. A comprehensive theory of such collections is contained in Moore [1], to which the reader is referred for proofs of the facts listed below. Each point P of Σ is on precisely one continuum $\Gamma \epsilon K$. Let us write $T(P) = \Gamma$ for the transformation which carries each point P of Σ into that continuum $\Gamma \epsilon K$ which contains P. The class K can be topologized in such a way that this transformation $T(P) = \Gamma$ becomes continuous. K is then again a Peano space. For greater clarity, we shall use Σ^* to denote the space that is obtained in this manner from K, and we shall use P^* to refer to a generic point of Σ^*.

1.14. The transformation T, defined in 1.13, may now be written in the form

$$T(P) = P^*, \qquad P \epsilon \Sigma, \qquad P^* \epsilon \Sigma^*.$$

$T(P)$ is single-valued and continuous on Σ. The inverse transformation $T^{-1}(P^*)$ need not be single-valued on Σ^*, but for every point P^* of Σ^* the inverse set $T^{-1}(P^*)$ is a continuum, namely one of the continua $\Gamma \epsilon K$. The following statements are consequences of the preceding statements (see Moore [1]).

a) If C^* is a continuum on Σ^*, then $T^{-1}(C^*)$ is a continuum on Σ.

b) If D^* is a domain (connected open set) on Σ^*, then $T^{-1}(D^*)$ is a domain on Σ.

c) If F^* is a closed set on Σ^*, then the set $\Sigma^* - F^*$ has the same number of components as the set $\Sigma - T^{-1}(F^*)$. In fact, the components of $\Sigma - T^{-1}(F^*)$ are simply the inverse sets of the components of $\Sigma^* - F^*$.

1.15. By means of the transformation $T(P) = P^*$, the continuous vector-function $\mathfrak{x}(P)$ is carried into a continuous vector-function $\mathfrak{x}^*(P^*)$ in the following sense. Given $P^* \epsilon \Sigma^*$, let us put

$$\mathfrak{x}^*(P^*) = \mathfrak{x}(P), \qquad P \epsilon T^{-1}(P^*). \tag{8}$$

Since $\mathfrak{x}(P)$ is constant on $T^{-1}(P^*)$, the relation (8) defines a single-valued and obviously continuous vector function on Σ^*. In a similar manner, the components $x_i(P)$ of $\mathfrak{x}(P)$, as well as the associated functions $f_i(P)$ defined in 1.4, are transformed by T into single-valued continuous functions on Σ^* which we shall denote by $x_i^*(P^*)$, $f_i^*(P^*)$ respectively. In a general way, if $g(P)$ is any continuous function on Σ, such that $g(P)$ is constant on each continuum $\Gamma \epsilon K$, then the formula

$$g^*(P^*) = g(P), \qquad P \epsilon T^{-1}(P^*)$$

transforms $g(P)$ into a single-valued continuous function $g^*(P^*)$ on Σ^*.

1.16. We shall call Σ^* *the associated Peano space* and the functions $\mathfrak{x}^*(P^*)$, $x_i^*(P^*)$, $f_i^*(P^*)$, $g^*(P^*)$ *the associated functions*, with respect to $\mathfrak{x}(P)$.

1.17. LEMMA. *Let $\mathfrak{x}(P)$ be a continuous vector function on a Peano space Σ. Let $f^*(P^*)$ be a continuous complex-valued function on the associated Peano space Σ^*, and let C^* be a continuum on Σ^*, such that $f^*(P^*)$ fails to satisfy condition* (Arg) *on C^*. Then the function*

$$f(P) = f^*(T(P))$$

fails to satisfy condition (Arg) *on the continuum $C = T^{-1}(C^*)$.*

PROOF. Obviously $f(P)$ is a single-valued continuous function on Σ, while $C = T^{-1}(C^*)$ is a continuum by 1.14. Assume now, in contradiction to the lemma, that $f(P)$ satisfies condition (Arg) on C. Let then ζ be any complex constant such that

$$f^*(P^*) - \zeta \neq 0 \quad \text{on} \quad C^*. \tag{9}$$

We have then also

$$f(P) - \zeta \neq 0 \quad \text{on} \quad C.$$

By assumption, we have therefore a single-valued continuous argument $\varphi(P)$ for $f(P) - \zeta$ on C. We assert that the formula

$$\varphi^*(P^*) = \varphi(P), \qquad P \in T^{-1}(P^*), \qquad P^* \in C^*,$$

defines a single-valued continuous argument for $f^*(P^*)$ on C^*. Clearly, we only have to show that

$$\varphi(P') = \varphi(P'')$$

for any two points P', P'' such that

$$P' \in T^{-1}(P^*), \qquad P'' \in T^{-1}(P^*), \qquad P^* \in C^*.$$

Now $T^{-1}(P^*)$ is a continuum Γ comprised in C, such that $f(P)$ is constant on Γ. Hence, $\varphi(P)$ is defined and continuous on $T^{-1}(P^*)$. Furthermore, for each $P \in T^{-1}(P^*)$, $\varphi(P)$ is an argument of the same complex number $f(P) = f^*(T(P)) = f^*(P^*)$. Hence the set of values of $\varphi(P)$ on $T^{-1}(P^*)$ is denumerable. As $T^{-1}(P^*)$ is a continuum and $\varphi(P)$ is continuous on $T^{-1}(P^*)$, it follows that $\varphi(P)$ is constant on $T^{-1}(P^*)$. Hence $\varphi(P') = \varphi(P'')$ for $P', P'' \in T^{-1}(P^*)$, $P^* \in C^*$. Since ζ was arbitrary, except for the condition (9), we reached a contradiction to the assumption that $f^*(P^*)$ fails to satisfy condition (Arg) on C^*, and the lemma is proved.

1.18. LEMMA. *Let $\mathfrak{x}(P)$ be a continuous vector-function on a Peano space Σ. Let $\mathfrak{x}^*(P^*)$ be the associated vector-function on the associated Peano space Σ^*. If there exists an indicator continuum for $\mathfrak{x}^*(P^*)$ on Σ^*, then there also exists an indicator continuum for $\mathfrak{x}(P)$ on Σ.*

PROOF. If C^* is an indicator continuum for $\mathfrak{x}^*(P^*)$ on Σ^*, then $C = T^{-1}(C^*)$ is an indicator continuum for $\mathfrak{x}(P)$ on Σ, as a direct consequence of 1.17.

1.19. A Peano space is a *dendrite* if it possesses no proper cyclic element. The following lemma is an easy consequence of the general theorem concerning

the structure of a Peano space in terms of cyclic chains (see Kuratowski-Whyburn [1]). Hence the details of the proof will be left to the reader.

LEMMA. *Suppose the Peano space Σ is a dendrite. Then for every $\delta > 0$ there exists on Σ a finite system of points $Q_1, Q_2, \cdots, Q_n$ with the following properties.*

1) *For $k = 1, 2, \cdots, n$, the set $\Sigma - Q_k$ has precisely two components.*

2) *Each component of the set $\Sigma - (Q_1 + \cdots + Q_n)$ has a diameter less than δ.*

1.20. If $g(P)$ is a real or complex-valued function on a Peano space Σ, and E is any subset of Σ, then $\omega(g(P); E)$ will denote the oscillation of $g(P)$ on E. That is

$$\omega(g(P); E) = \text{l.u.b.} \mid g(P'') - g(P') \mid, \qquad P', P'' \in E.$$

Similarly, if $\mathfrak{x}(P)$ is a continuous vector-function on Σ, its oscillation on E is given by

$$\omega(\mathfrak{x}(P); E) = \text{l.u.b.} \mid \mathfrak{x}(P'') - \mathfrak{x}(P') \mid, \qquad P', P'' \in E,$$

where $\mid \mathfrak{x}(P'') - \mathfrak{x}(P') \mid$ designates the length of the vector $\mathfrak{x}(P'') - \mathfrak{x}(P')$.

1.21. The following elementary lemma, due to Geöcze, is very important both in his work and in ours.

LEMMA. *Given a continuous path-surface as in* (3), *suppose that for every $\epsilon > 0$ there exists on U a finite system of simple closed curves $C_1, C_2, \cdots, C_m$ with the following properties.* (*i*) *No two of these curves intersect each other.* (*ii*) *On each of the regions into which U is divided by this system of curves, the oscillation of $\mathfrak{x}(P)$ is less than ϵ. Then $A(S) = 0$.*

PROOF. From (i) it follows that the curves C_j determine on U exactly $m + 1$ regions which we denote, in any order, by $R_1, R_2, \cdots, R_{m+1}$. From (ii) it follows that the oscillation of $\mathfrak{x}(P)$ on each of the curves C_j is also $\leqq \epsilon$. Let us now select on each one of the curves C_j a point P_j, and in each one of the regions R_k an interior point Q_k. We have then

$$(10) \qquad \mid \mathfrak{x}(P) - \mathfrak{x}(P_j) \mid \leqq \epsilon \quad \text{for} \quad P \in C_j, \qquad j = 1, 2, \cdots, m,$$

$$(11) \qquad \mid \mathfrak{x}(Q) - \mathfrak{x}(Q_k) \mid \leqq \epsilon \quad \text{for} \quad Q \in R_k, \qquad k = 1, 2, \cdots, m + 1.$$

Let us now take a triangulation T of U, such that a) the maximum diameter of the triangles of T is less than ϵ, b) each one of the curves C_j is a sum of sides of triangles of T, and c) no triangle of T has vertices on two different curves C_j. Such a triangulation T obviously exists. We now proceed to define a corresponding polyhedron $\mathfrak{P}$ (see 0.4) as follows. Let V be a vertex of T. We assign the point V' as follows. If V is an interior point of a region R_k, then $V' = \mathfrak{x}(Q_k)$. If V is on a curve C_j, then $V' = \mathfrak{x}(P_j)$. In view of (10), (11) and condition a), it follows that the resulting polyhedron $\mathfrak{P}$ satisfies the condition

$$(12) \qquad d(S, \mathfrak{P}) \leqq 2\epsilon.$$

Further, since each triangle of T has either two vertices on the same curve C_j,

or else has at least two vertices interior to the same region R_k, it follows that

$$E(\mathfrak{P}) = 0. \tag{13}$$

Thus, for every $\epsilon > 0$, we have a polyhedron $\mathfrak{P}$ satisfying (12) and (13). By 0.6 it follows that $A(S) = 0$.

1.22. LEMMA. *Given a continuous path-surface S as in* (3), *suppose that there exists an indicator curve for $\mathfrak{x}(P)$ on U. Then $A(S) > 0$.*

This lemma is well known, except for the terminology used here. See for example Radó [1]; the proof given there for Jordan surfaces applies, after obvious modifications, to a general continuous path-surface. In verifying the lemma, the term indicator curve should be used in the equivalent sense described in 1.9.

CHAPTER II. THE MAIN RESULTS

2.1. The fundamental lemma. *Let Σ be a Peano space, such that for every totally disconnected closed set Δ on Σ, the set $\Sigma - \Delta$ is connected. Let the continuous vector-function $\mathfrak{x}(P)$ be non-degenerate on Σ. Then there exists an indicator curve for $\mathfrak{x}(P)$ on Σ.*

PROOF. In view of 1.12, it is sufficient to prove that there exists an indicator continuum for $\mathfrak{x}(P)$ on Σ. Since $\mathfrak{x}(P)$ is non-degenerate, one of its components, say $x_1(P)$, is not constant on Σ. Let m_1, M_1 be the minimum and the maximum of $x_1(P)$ on Σ respectively. We have then $m_1 < M_1$, and there exist two points O, O^* on Σ such that

$$x_1(O) = m_1, \qquad x_1(O^*) = M_1.$$

Put

$$\xi = \frac{m_1 + M_1}{2},$$

and consider on Σ the set

$$E = \underset{P}{E}[x_1(P) < \xi].$$

That is, E consists of those points P of Σ for which $x_1(P) < \xi$. Then E is a non-vacuous open set, and we have

$$O \in E, \qquad O^* \in \Sigma - \bar{E}$$

where $\bar{E}$ designates the closure of E. Let D be the component of E which contains O. Then D is a domain containing O. The frontier $\mathcal{F}(D)$ of D is a closed set which separates O and O^*. Hence, by our assumption concerning the space Σ, $\mathcal{F}(D)$ is not totally disconnected. Hence $\mathcal{F}(D)$ has some component C_0 which is a continuum not reducing to a single point. Since $\mathfrak{x}(P)$ is non-degenerate by assumption, and since $x_1(P) \equiv \xi$ on C_0, at least one of $x_2(P)$, $x_3(P)$, say $x_2(P)$, is not constant on C_0. Since C_0 is compact, we have therefore on C_0 two points Q_0, Q_0^* such that

$$x_2(Q_0) = m_2, \qquad x_2(Q_0^*) = M_2,$$

where m_2, M_2 designate the minimum and the maximum respectively of $x_2(P)$ on C_0. Note that

$$m_2 < M_2,$$

since $x_2(P)$ is not constant on C_0.

Since Σ is locally connected and $\mathfrak{x}(P)$ is continuous, we have two continua κ_0, κ_0^*, containing Q_0, Q_0^* as interior points respectively, and so small in diameter that

$$\omega(x_1(P), \kappa_0), \qquad \omega(x_1(P), \kappa_0^*) < \frac{M_2 - m_2}{4},$$

$$\omega(x_2(P), \kappa_0), \qquad \omega(x_2(P), \kappa_0^*) < \frac{M_2 - m_2}{4}.$$

As Q_0, Q_0^* are points of the frontier $\mathcal{F}(D)$ of D, there exist points A_0, A_0^* such that

$$A_0 \in \kappa_0 \cdot D, \qquad A_0^* \in \kappa_0^* \cdot D.$$

Let γ be an arc that joins A_0 and A_0^* in D. As γ is a closed subset of D and $x_1(P) < \xi$ on D, we have a number ξ_0 such that

$$x_1(P) \leqq \xi_0 < \xi \quad \text{for} \quad P \in \gamma.$$

Let us put

$$\zeta_3 = \frac{\xi_0 + \xi}{2} + i\frac{M_2 + m_2}{2}.$$

2.2. We shall use the function (cf. 1.4)

$$f_3(P) = x_1(P) + ix_2(P)$$

to show presently that the set

$$\Gamma = C_0 + \kappa_0 + \kappa_0^* + \gamma$$

is an indicator continuum for $\mathfrak{x}(P)$ on Σ. This will be achieved by establishing the following facts.

a) Γ is a continuum.

b) $f_3(P) - \zeta_3 \neq 0$ on Γ.

c) $f_3(P) - \zeta_3$ does not possess a single-valued continuous argument on Γ.

2.3. Now (a) in 2.2 is obvious, since each one of the sets C_0, κ_0, κ_0^*, γ is a continuum, and

$$C_0\kappa_0 \neq 0, \qquad \kappa_0\gamma \neq 0, \qquad \gamma\kappa_0^* \neq 0, \qquad \kappa_0^*C_0 \neq 0.$$

2.4. Next we show that

$$\Re(f_3(P) - \zeta_3) > 0 \quad \text{on} \quad C_0, \tag{14}$$

$$\Im(f_3(P) - \zeta_3) < 0 \quad \text{on} \quad \kappa_0, \tag{15}$$

(16) $$\mathfrak{I}(f_3(P) - \zeta_3) > 0 \quad \text{on} \quad \kappa_0^*,$$

(17) $$\mathfrak{R}(f_3(P) - \zeta_3) < 0 \quad \text{on} \quad \gamma.$$

In these inequalities, $\mathfrak{R}$ and $\mathfrak{I}$ denote real part and imaginary part respectively, as usual. To verify these relations, observe that we have on C_0

$$\mathfrak{R}(f_3(p) - \zeta_3) = \xi - \frac{\xi_0 + \xi}{2} = \frac{\xi - \xi_0}{2} > 0.$$

On κ_0 we have

$$\begin{aligned}\mathfrak{I}(f_3(P) - \zeta_3) &= x_2(P) - \frac{M_2 + m_2}{2} \\ &\leqq x_2(Q_0) + \omega(x_2(P); \kappa_0) - \frac{M_2 + m_2}{2} \\ &\leqq m_2 + \frac{M_2 - m_2}{4} - \frac{M_2 + m_2}{2} = \frac{m_2 - M_2}{4} < 0.\end{aligned}$$

On κ_0^* we have

$$\begin{aligned}\mathfrak{I}(f_3(P) - \zeta_3) &= x_2(P) - \frac{M_2 + m_2}{2} \\ &\geqq x_2(Q_0^*) - \omega(x_2(P); \kappa_0^*) - \frac{M_2 + m_2}{2} \\ &\geqq M_2 - \frac{M_2 - m_2}{4} - \frac{M_2 + m_2}{2} = \frac{M_2 - m_2}{4} > 0.\end{aligned}$$

On γ we have

$$\begin{aligned}\mathfrak{R}(f_3(P) - \zeta_3) &= x_1(P) - \frac{\xi_0 + \xi}{2} \\ &\leqq \xi_0 - \frac{\xi_0 + \xi}{2} = \frac{\xi_0 - \xi}{2} < 0.\end{aligned}$$

2.5. Condition (b) in 2.2 is an immediate consequence of (14), (15), (16), (17) in 2.4.

2.6. Let us note the following inequalities.

(18) $$\mathfrak{R}(f_3(Q_0) - \zeta_3) > 0,$$

(19) $$\mathfrak{I}(f_3(Q_0) - \zeta_3) < 0,$$

(20) $$\mathfrak{R}(f_3(Q_0^*) - \zeta_3) > 0,$$

(21) $$\mathfrak{I}(f_3(Q_0^*) - \zeta_3) > 0,$$

(18) and (20) follow directly from (14), while (19), (21) follow from (15) and (16) respectively.

2.7. Assume now that $f_3(P) - \zeta_0$ possesses a single-valued continuous argument $\varphi(P)$ on Γ, in contradiction to our assertion (c) in 2.2. By (18), (19) we can assume then that

$$-\frac{\pi}{2} < \varphi(Q_0) < 0. \tag{22}$$

Consider now $\varphi(P)$ on the continuum C_0. We assert that

$$-\pi < \varphi(P) < \pi \quad \text{on} \quad C_0 .$$

Otherwise, by virtue of its continuity and in view of (22), $\varphi(P)$ would have to assume at some point of C_0 one of the values $-\pi$, π, which is however impossible by (14). In particular it follows that

$$-\pi < \varphi(Q_0^*) < \pi. \tag{23}$$

In view of (20), (21) it follows from (23) that

$$0 < \varphi(Q_0^*) < \frac{\pi}{2}. \tag{24}$$

Consider now $\varphi(P)$ on the continuum $\kappa_0 + \gamma + \kappa_0^*$. We assert that

$$-2\pi < \varphi(P) < 0 \quad \text{on} \quad \kappa_0 + \gamma + \kappa_0^* .$$

Otherwise, by virtue of its continuity and in view of (22), $\varphi(P)$ would have to assume on $\kappa_0 + \gamma + \kappa_0^*$ one of the values 0, -2π, which is however impossible by (15), (16), (17). In particular, it follows that

$$-2\pi < \varphi(Q_0^*) < 0. \tag{25}$$

Since (24) and (25) obviously contradict each other, the assertion (c) in 2.2 is established, and the proof of the fundamental lemma in 2.1 is complete.

2.8. THEOREM (cf. 0.7). *Let S be a continuous path-surface given as in* (3). *If $\mathfrak{x}(P)$ is non-degenerate on U, then $A(S) > 0$.*

PROOF. It is well known that U is not disconnected by any one of its totally disconnected closed subsets. Hence, by 2.1, there exists an indicator curve for $\mathfrak{x}(P)$ on U. By 1.22 it follows that $A(S) > 0$.

2.9. THEOREM (of Geöcze). *If the representation* (3) *defines a bi-unique correspondence between the points P of U and the points (x_1, x_2, x_3) of S, then $A(S) > 0$.* Briefly: *the Lebesgue area of a simple continuous path-surface is always positive.* This theorem is of course a special case of 2.8.

2.10. LEMMA. *Let S be a continuous path-surface given as in* (3). *Suppose there exists no indicator curve for $\mathfrak{x}(P)$ on U. Then the associated Peano space Σ^* reduces to a dendrite.*

PROOF. By a well-known theorem of R. L. Moore [2], Σ^* is a *cactoid*. That is, either Σ^* is a dendrite, or else every proper cyclic element of Σ^* is homeomorphic to the surface of the unit sphere. Consider now the associated vector-function $\mathfrak{x}^*(P^*)$ defined in 1.16. By 1.12, there exists no indicator continuum

for $\mathfrak{x}(P)$ on U. By 1.18, it follows that there exists no indicator continuum for $\mathfrak{x}^*(P^*)$ on Σ^*. Since $\mathfrak{x}^*(P^*)$ is non-degenerate on Σ^*, and since by the theorem of R. L. Moore, referred to above, no proper cyclic element of Σ^* is disconnected by any one of its closed totally disconnected subsets, it follows from the fundamental lemma in 2.1, applied to an hypothetical proper cyclic element of Σ^*, that Σ^* cannot have proper cyclic elements, and therefore reduces to a dendrite, as asserted.

2.11. Lemma. Given a continuous path-surface S as in (3), suppose that the associated Peano space Σ^* reduces to a dendrite. Then for every $\epsilon > 0$ there exists on U a finite system of simple closed curves $C_1, C_2, \cdots, C_m$ with the following properties. (i) No two of the curves $C_1, C_2, \cdots, C_m$ intersect each other. (ii) On each component of the set $U - (C_1 + C_2 + \cdots + C_m)$ the oscillation of $\mathfrak{x}(P)$ is less than ϵ.

Proof. Consider again the associated vector-function $\mathfrak{x}^*(P^*)$ on Σ^*. By 1.19, for every $\delta > 0$ we have on Σ^* a finite system of points $P_1^*, P_2^*, \cdots, P_n^*$ with the following properties. (i) For each k, the set $\Sigma^* - P_k^*$ has precisely two components. (ii) Each component of the set $\Sigma^* - (P_1^* + P_2^* + \cdots + P_n^*)$ has a diameter less than δ. Since $\mathfrak{x}^*(P^*)$ is uniformly continuous on Σ^*, we can choose δ so small that the oscillation of $\mathfrak{x}^*(P^*)$ on each component of the set $\Sigma^* - (P_1^* + P_2^* + \cdots + P_n^*)$ is less than ϵ. Let $\Gamma_1, \Gamma_2, \cdots, \Gamma_n$ be the continua which correspond to $P_1^*, P_2^*, \cdots, P_n^*$ on U. Then (cf. 1.14) for each k the set $U - \Gamma_k$ has precisely two components, and on each component of the set $U - (\Gamma_1 + \cdots + \Gamma_n)$ the oscillation of $\mathfrak{x}(P)$ is less than ϵ. By well-known theorems on the topology of the sphere (see for example Newman [1]), the proof can now be completed as follows. The set $U - (\Gamma_1 + \cdots + \Gamma_n)$ has precisely $n + 1$ components, each of which is a domain (connected open set). Let $D_1, D_2, \cdots, D_{n+1}$ be these domains, in any order. Consider one of these domains D_k. There exists in D_k a sequence of Jordan regions R_k^j, $j = 1, 2, \cdots$, each bounded by a finite number of simple closed curves, such that (i) $R_k^j \subset R_k^{j+1}$, and (ii) for every point $P \in D_k$ there exists a j_0 such that $P \in R_k^j$ for $j > j_0$. For fixed large j, U is then the sum of $2n + 1$ Jordan regions, without common interior points, namely of the $(n + 1)$ regions $R_1^j, \cdots, R_n^j, R_{n+1}^j$ and of n further doubly connected Jordan regions $r_1^j, \cdots, r_n^j$, each of which contains precisely one of the continua $\Gamma_1, \cdots, \Gamma_n$ in its interior. These latter regions $r_1^j, \cdots, r_n^j$ are doubly connected as a consequence of the fact that for each k the set $U - \Gamma_k$ has precisely two components. Each r_k^j is bounded by two simple closed curves which are parts of the boundaries of two different regions R_k^j. The oscillation of $\mathfrak{x}(P)$ on R_k^j is less than ϵ, since $R_k^j \subset D_k$. On the other hand, if E_k is any open set containing Γ_k, we shall have $r_k^j \subset E_k$ for j sufficiently large. Since $\mathfrak{x}(P)$ is constant on Γ_k, the open set E_k can be so chosen that the oscillation of $\mathfrak{x}(P)$ on E_k is less than ϵ. Hence, if j is sufficiently large, the system consisting of the boundary curves of all the regions $R_1^j, R_2^j, \cdots, R_{n+1}^j$ will possess the properties required in the lemma.

2.12. The main theorem. *Given a continuous path-surface S as in* (3), *the*

Lebesgue area $A(S)$ of S is equal to zero if and only if there exists no indicator curve for $\mathfrak{x}(P)$ on U.

PROOF. The condition is necessary by 1.22. To prove the sufficiency, assume that there exists no indicator curve for $\mathfrak{x}(P)$ on U. By 2.10, the associated Peano space reduces then to a dendrite. By 2.11, it follows that the assumptions of the lemma in 1.21 are satisfied, and hence $A(S) = 0$.

2.13. THEOREM (cf. 0.7). *The Lebesgue area $A(S)$ of a continuous path-surface S, given as in (3), is equal to zero if and only if the associated Peano space Σ^* reduces to a dendrite.*

PROOF. The necessity follows directly from 1.22 and 2.10, while the sufficiency follows from 2.11 and 1.21.

2.14. THEOREM (of Geöcze). *The Lebesgue area $A(S)$ of a continuous path-surface S, given as in (3), is equal to zero if and only if for every $\epsilon > 0$ there exists on U a finite system of simple closed curves $C_1, C_2, \cdots, C_m$ with the following properties. (i) No two of these curves intersect each other. (ii) The oscillation of $\mathfrak{x}(P)$ is less than ϵ on each component of the set $U - (C_1 + C_2 + \cdots + C_m)$.*

PROOF. The necessity follows from 2.13 and 2.11, while the sufficiency follows from 1.21.

CHAPTER III. MISCELLANEOUS REMARKS

3.1. Given a continuous path-surface S as in (3), let us denote by $[S]$ the set of points (x_1, x_2, x_3) which correspond to the points P of U by means of (3). We already observed that the area $A(S)$ of S is not determined by the point-set $[S]$, but rather by the manner in which $[S]$ is described by virtue of (3). This remark will be illustrated presently by examples, due essentially to Geöcze.

3.2. Let us denote by M_0 the half-circle on U which passes through the points $(0, 0, -1)$, $(1, 0, 0)$, $(0, 0, 1)$. Let $\mathfrak{x}_0(P)$ be any continuous vector-function defined on M_0. If P is a point of U not on M_0, then let us denote by $\pi(P)$ the small circle on U which passes through P and is contained in a plane parallel to the (x_1, x_2) plane. Let the vector-function $\mathfrak{x}(P)$ be defined on U as follows.

$$\mathfrak{x}(P) = \mathfrak{x}_0(P) \text{ on } M_0,$$

$$\mathfrak{x}(P) = \text{constant on } \pi(P).$$

Then $\mathfrak{x}(P)$ is clearly continuous on U. Consider now the continuous path-surface S determined by this particular vector-function $\mathfrak{x}(P)$. We assert that $A(S) = 0$. Indeed, let T be any triangulation of U, such that each triangle of T has a side which is a sub-arc of a great circle passing through the points $(0, 0, 1)$ and $(0, 0, -1)$. For each vertex V of T, let us assign the point $\mathfrak{x}(V)$ as the point V' of 0.4. We obtain in this manner a polyhedron $\mathfrak{P}$, such that $d(S, \mathfrak{P})$ is equal to the maximum diameter of the triangles of T, and such that $E(\mathfrak{P}) = 0$. By 0.6, it follows that $A(S) = 0$. This was to be expected, since the path-surface S under consideration may be thought of as reducing, in a

way, to the continuous path-curve Γ given by

$$\Gamma: \quad \mathfrak{x} = \mathfrak{x}_0(P), \qquad P \in M_0. \tag{26}$$

Let us designate by $[\Gamma]$ the set of those points (x_1, x_2, x_3) which correspond to the points P of M_0 by means of (26). Clearly then $[\Gamma] = [S]$. On the other hand, it is well known that by proper choice of $\mathfrak{x}_0(P)$ the point-set $[\Gamma]$ and hence also the point-set $[S]$ can be made to coincide with any desired compact, connected and locally connected point-set in (x_1, x_2, x_3) space.

3.3. In particular, we can make $[S]$ coincide with the cube $-1 \leqq x_i \leqq 1$, $i = 1, 2, 3$. We obtain then the continuous path-surface S of Geöcze [3] which fills a cube and has zero area.

3.4. Next, let us choose $[S]$ as a point-set U' homeomorphic to U, the surface of the unit sphere. The construction described in 3.2 yields then a continuous path-surface S such that $[S] = U'$ and $A(S) = 0$. On the other hand, we have by assumption a continuous vector-function $\bar{\mathfrak{x}}(P)$ on U such that $\bar{\mathfrak{x}}(P_1) \neq \bar{\mathfrak{x}}(P_2)$ for $P_1 \neq P_2$, and such that for the continuous path-surface

$$\bar{S}: \quad \mathfrak{x} = \bar{\mathfrak{x}}(P), \qquad P \in U$$

we have $[\bar{S}] = U'$. By 2.9 we have then $A(\bar{S}) > 0$. Thus we have two continuous path-surfaces

$$S: \quad \mathfrak{x} = \mathfrak{x}(P), \qquad P \in U, \tag{27}$$

$$\bar{S}: \quad \mathfrak{x} = \bar{\mathfrak{x}}(P), \qquad P \in U, \tag{28}$$

with the following properties. (i) $[S] = [\bar{S}] = U'$, where U' is homeomorphic to U. (ii) For each point (x_1, x_2, x_3) on U', we have *precisely one point* $P \in U$ which is carried into (x_1, x_2, x_3) by (28) and we have *at least one point* $P \in U$ which is carried into (x_1, x_2, x_3) by (27). (iii) Finally, contrarily to what one may have expected under such circumstances, we have $A(S) < A(\bar{S})$, since $A(S) = 0$, $A(\bar{S}) > 0$.

3.5. In a more picturesque way, we can describe this situation as follows. Let U' be a topological image of the surface U of the unit sphere. Then the area of a continuous path-surface S, given as in (3), is certainly positive if S covers U' *just once*. On the other hand, if S is required to cover all of U', but is permitted to cover *parts of U' several times*, then the area of S may be equal to zero. Of course, there is nothing really paradoxical about this, but such a situation is an illustration of the fundamental differences between the theory of the length of curves and the theory of the area of surfaces.

3.6. We pass on to remarks concerning certain interesting aspects of our main theorem in 2.12. A *closed continuous path-curve* Γ, in (x_1, x_2, x_3) space, is determined by an equation

$$\Gamma: \quad \mathfrak{x} = \mathfrak{x}(P), \qquad P \in C, \tag{29}$$

where C is a simple closed curve and $\mathfrak{x}(P)$ is a continuous vector-function on C. It is assumed that a positive direction is assigned on C. Let p be any plane.

Let P' be the point (x_1, x_2, x_3) which corresponds to the point P of C by means of (29), and let P_p be the orthogonal projection of P' upon the plane p. Let finally $\mathfrak{x}_p(P)$ be the vector that joins the point $(0, 0, 0)$ to the point P_p. Then we shall designate by Γ_p the closed continuous path-curve

$$\Gamma_p: \qquad \mathfrak{x} = \mathfrak{x}_p(P), \qquad P \in C.$$

We shall call Γ_p *the orthogonal projection* of Γ upon the plane p. Let Q_p be a variable point in the plane p. We define then the index-function $\mu(Q_p, \Gamma_p)$ as follows. If Q_p is on Γ_p, then $\mu(Q_p, \Gamma_p) = 0$. Otherwise, $\mu(Q_p, \Gamma_p)$ is equal to the topological index of Q_p with respect to Γ_p. For conciseness, we shall use the subscripts 1, 2, 3 to refer to the coordinate planes (x_2, x_3), (x_3, x_1), (x_1, x_2) respectively. Then the index-functions $\mu(Q_i, \Gamma_i)$, $i = 1, 2, 3$, are identical with the index-functions $\mu_i(\zeta_i)$ of 1.9.

3.7. Consider now a continuous path-surface

$$(30) \qquad S: \qquad \mathfrak{x} = \mathfrak{x}(P), \qquad P \in U.$$

If C is any simple closed curve on U, then there corresponds to C, by means of (30), a closed continuous path-curve

$$(31) \qquad \Gamma: \qquad \mathfrak{x} = \mathfrak{x}(P), \qquad P \in C.$$

Suppose now that there exists no indicator curve for $\mathfrak{x}(P)$ on U. By our main theorem in 2.12, $A(S) = 0$. Since $A(S)$ is invariable under changes of the Cartesian coordinate system, it follows (cf. 1.22, 1.9, 3.6) that $\mu(Q_p, \Gamma_p) \equiv 0$ for every choice of the plane p. Hence we have the

THEOREM. *If for every closed continuous path-curve Γ on a continuous path-surface S, given as in* (30) *and* (31) *respectively, the three index-functions $\mu_i(\zeta_i)$, $i = 1, 2, 3$, are identically equal to zero, then the index-function $\mu(Q_p, \Gamma_p)$ is also identically equal to zero for every choice of the plane p.*

3.8. One may be tempted to strengthen the preceding statement as follows: If for a closed continuous path-curve Γ, given as in (29), the three index-functions $\mu(Q_i, \Gamma_i)$, $i = 1, 2, 3$, are identically equal to zero, then the index-function $\mu(Q_p, \Gamma_p)$ is also identically equal to zero, for every choice of the plane p. *This statement is however false.* A first counter-example has been suggested to the writer by J. v. Neumann. Subsequently, W. Scott, a student of the writer, found an extremely simple construction for examples serving various further purposes also.

3.9. In conclusion, we shall give a brief description of the condition used by Geöcze in his characterization of continuous path-surfaces of zero area. Geöcze assumes that the vector-function $\mathfrak{x}(P)$ which determines the surface is defined on the unit square. Hence, his continuous path-surface is given by an equation

$$(32) \qquad S: \qquad \mathfrak{x} = \mathfrak{x}(u, v), \qquad (u, v) \in s_0,$$

where

$$(33) \qquad s_0: \qquad 0 \leqq u \leqq 1, \qquad 0 \leqq v \leqq 1.$$

In stating the definitions of Geöcze, we shall use present-day terminology if possible; otherwise, we shall use literal translations of the terms invented by Geöcze. All point-sets occurring in 3.10 to 3.15 are in the (u, v) plane unless stated otherwise.

3.10. A *chain* consists of a finite number of straight segments placed end to end. These segments are permitted to intersect each other.

3.11. We recall that a domain D is a connected open set. The frontier $\mathfrak{F}(D)$ of D satisfies the relation $\mathfrak{F}(D) = \bar{D} - D$, where $\bar{D}$ is the closure of D.

3.12. If D is a simply connected domain, then I will denote the set of those points of $\mathfrak{F}(D)$ which are accessible from within D by some chain. The points of I can be cyclicly ordered.

3.13. A domain D is *simple* if the following conditions hold. (i) D is bounded. (ii) Let D_∞ denote the unbounded component of the complement of $\bar{D}$. Then $\mathfrak{F}(D) = \mathfrak{F}(D_\infty)$. Geöcze notes that a simple domain is also simply connected.

3.14. A *four-sided domain* is a figure consisting of a *simple* domain (see 3.13) and of four point-sets $\alpha_1, \alpha_2, \alpha_3, \alpha_4$, called the *sides*, satisfying the following conditions. (i) $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ are subsets of the set I defined in 3.12. (ii) $\bar{\alpha}_1\bar{\alpha}_3 = 0$, $\bar{\alpha}_2\bar{\alpha}_4 = 0$, where the bars denote closure. (iii) $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = I$. (iv) If P_1, P_2, P_3, P_4 are any four points such that $P_i \in \alpha_i$, $i = 1, 2, 3, 4$, then the cyclic order of these points, on the set I, is P_1, P_2, P_3, P_4.

3.15. Given the continuous path-surface (32), the symbol $(\Phi, \Psi, \lambda_1, \lambda_2, \lambda_3, \lambda_4)$ will be meaningful if the following conditions are satisfied. (i) Φ and Ψ are two different ones of the components $x_1(u, v)$, $x_2(u, v)$, $x_3(u, v)$, of the vector-function $\mathfrak{x}(u, v)$ appearing in (32). (ii) $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are real numbers such that $\lambda_3 > \lambda_1$, $\lambda_4 > \lambda_2$. (iii) There exists in the unit square s_0 a *four-sided domain*, with sides $\alpha_1, \alpha_2, \alpha_3, \alpha_4$, such that $\Phi = \lambda_1$ on α_1, $\Phi = \lambda_3$ on α_3, $\Psi = \lambda_2$ on α_2, $\Psi = \lambda_4$ on α_4. Such a four-sided domain will then be denoted by the symbol $(\Phi, \Psi, \lambda_1, \lambda_2, \lambda_3, \lambda_4)$.

3.16. THEOREM OF GEÖCZE (see Geöcze [4]). *The Lebesgue area $A(S)$ of the continuous path-surface S, given by* (32), *is equal to zero if and only if there exists no four-sided domain* $(\Phi, \Psi, \lambda_1, \lambda_2, \lambda_3, \lambda_4)$.

3.17. It is easy to see that the existence of a domain $(\Phi, \Psi, \lambda_1, \lambda_2, \lambda_3, \lambda_4)$ implies the existence of an indicator curve in our sense (see 1.8, 1.9; in the definitions given there, U should of course be replaced now by s_0). But the present writer was unable to derive a proof of reasonable directness and simplicity for the converse statement and hence for the equivalence of the Geöcze condition (see 3.15) with our condition (see 2.12).

THE OHIO STATE UNIVERSITY

BIBLIOGRAPHY

EILENBERG, S. [1] *Transformations continues en circonférences et la topologie du plan,* Fundamenta Mathematicae, vol. 26, 1936, pp. 61–112.

GEÖCZE, Z. [1] *On the theory of the area* (in Hungarian), Mathematikai és természettudományi értesitő, vol. 31, 1913, pp. 306–318.

——— [2] *On the surface of zero area* (in Hungarian), ibid. vol. 33,1915, pp. 730–748.

GEÖCZE, Z. [3] *Example of a surface of zero area which fills a cube* (in Hungarian), Mathematikai és physikai lapok, vol. 23, 1914, pp. 115–117.

Recherches générales sur la quadrature des surfaces courbes, Part 1, Mathematische und naturwissenschaftliche Berichte aus Ungarn, vol. 27, 1909, pp. 1–21.

KURATOWSKI, C., AND WHYBURN, G. T. [1] *Sur les éléments cyclic et leurs applications*, Fundamenta Mathematicae, vol. 16, 1930, pp. 305–331.

MOORE, R. L. [1] *Foundations of point-set theory*, American Mathematical Society Colloquium publications, vol. 13, 1932.

——— [2] *Concerning upper semi-continuous collections*, Monatshefte für Mathematik und Physik, vol. 36, 1929, pp. 81–88.

MORREY, C. B. [1] *An analytic characterization of surfaces of finite Lebesque area, Part 1*, American Journal of Mathematics, vol. 57, 1935, pp. 692–702.

NEWMAN, M. H. A. [1] *Elements of the topology of plane sets of points*, Cambridge University Press, 1939.

RADO, T. [1] *Sur l'aire des surfaces courbes*, Acta Szeged, vol. 3, 1927, pp. 131–169.

——— [2] *On the semi-continuity of double integrals in parametric form*, Transactions of the American Mathematical Society, vol. 51, 1942, pp. 336–361.

ANNALS OF MATHEMATICS
Vol. 44, No. 2, April, 1943

AUTOMORPHISM RINGS OF PRIMARY ABELIAN OPERATOR GROUPS[1]

BY REINHOLD BAER

(Revision received March 7, 1942)

The representation of a ring as the ring of all the automorphisms of a primary abelian operator group expresses significant inner properties of this ring, since there exists essentially at most one such representation of a given ring. Thus it is the object of this investigation to expose invariant qualities of these automorphism rings. They are found to be pecularities of their ideal theory (Chapter III); and from these we select a complete set of postulates for the class of all the automorphism rings of primary abelian operator groups (Sections **12** and **13**); thus proving incidentally that the structure of these rings is completely determined by their ideal theory (Section **14**).

The properties of the ideals in these rings reflect so completely the structure of the underlying operator group that we are able to prove—provided the "rank" of this group is at least three—the essential identity of the group of automorphisms of the ring and of the group of projectivities (= biunivoque and monotone increasing transformations) of the system of admissible subgroups of the underlying operator group (Chapter II).

Two [extreme] special cases may serve as an illustration for these theorems. Every projective geometry of finite dimension not less than three may be represented as the set of admissible subgroups of a suitable primary abelian operator group; and our characterization of the automorphisms of the automophism ring specializes in this case to the (well known) theorem that each automorphism of the automorphism ring (= ring of square matrices with coefficients from a suitable (not necessarily commutative) field) is induced by a so-called semi-linear transformation.

If secondly G is an ordinary primary abelian group with the property that the least common multiple of the orders of the elements in G is a prime power p^m and that G contains at least three independent elements of maximum order, then our theorem states that every automorphism of the automorphism ring is an inner auomorphism.

CHAPTER I: THE RELATIONS OF A GROUP TO ITS AUTOMORPHISM RING

1. The Admissible Subgroups

In this section we collect a number of definitions and facts concerning primary abelian operator groups which will be needed in the future.[2]

---

[1] Presented to the American Math. Soc. September, 1941.

[2] Proofs and details may be found in Baer (4). For references cp. the Bibliography at the end of the paper.

1A. *Cycles.* The element c in the partially ordered set D is termed a *cycle of order* $n = n(c)$, if exactly $n + 1$ elements in D are parts of c, and if the parts of c form an ordered set; if c is a cycle of order n, then it is possible to write down the system of all its parts in the following fashion: $0 = c^{(n)} < c^{(n-1)} < \cdots < c^{(2)} < c' < c^{(0)} = c$.

All the partially ordered sets considered will be complete lattices satisfying the modular (Dedekind's) law. The smallest element containing a certain set of elements will be termed its join and the greatest element contained in all the elements of some set will be called its meet or cross-cut. The following fact[3] is of great importance:

If the element s in the modular lattice D is the join of cycles whose orders do not exceed n, and if c is a subcycle of s, then the order of c does not exceed n.

The subset S of the complete modular lattice D is said to be *independent*, if the meet of the element s in S and of the join of the other elements in S is 0 for every element s in S. We shall usually impose the further restriction that the elements in an independent set should all be different from 0.

1B. *Rings of Subgroups.* Suppose that G is an abelian group and that the composition of its elements is written as addition. Then a system D of subgroups of G shall be termed a ring of subgroups, if it contains 0 and G, and if it contains the cross-cuts and sums[4] of all the subsets of D. It is well known that every ring of subgroups of G is a complete modular lattice; and it is obvious that a group G may contain sets of subgroups which are complete modular lattices, but not rings of subgroups.

1C. *Operator Groups.* Suppose that G is an abelian group and that E is a ring which contains an identity element 1. We say that G is an operator group over E, or that G admits the elements in E as operators, if there exists to every element g in G and to every element e in E an element ge in G such that $(g' \pm g'')e = g'e \pm g''e$, $g(e' \pm e'') = ge' \pm ge''$, $g(e'e'') = (ge')e'' = ge'e''$, $g1 = g$ for g, g', g'' in G, e, e', e'' in E. A subset S of G is an *E-admissible subgroup* if, and only if, $S = S + S$ and $S = SE$. If x is any element in G, then xE is the *E-admissible subgroup generated by* x.

The set of all the E-admissible subgroups in G is denoted by $D(G;E)$. Clearly $D(G;E)$ is a ring of subgroups of G. The set S of elements in G is said to be *independent*, if the elements in S generate different subgroups which are all different from 0, and if the subgroups generated by the elements in S form an independent set of elements in $D(G;E)$.

1D. *Primary Rings.* The ring E is said to be a primary ring, if it contains an identity element 1, and if there exists in E a two-sided ideal $P = P(E)$ such that

---

[3] Baer (4), Theorem I.2.1.

[4] Cross-cut signifies the set-theoretical cross-cut of the subgroups whereas sum is the group-theoretical sum of the subgroups.

$P^m = 0$ for some (smallest) positive $m = m(E)$ and such that every right- and every left-ideal in E is a power of P. Then there exist elements p in E such that $P = pE = Ep$ and such elements p we term *primes in E*. If p is a prime in E, then every ideal in E has the form $P^i = p^iE = Ep^i$. An element in E does not belong to P if, and only if, it possesses an inverse in E. We note finally that the four properties: $P = 0$, $p = 0$, $m = 1$, and "E is a field" are equivalent.[5]

1E. *Primary Abelian Operator Groups.* If the abelian group G admits the elements in the primary ring E as operators, then G is said to be a *primary abelian operator group* (*over E*). If x is an element in the primary abelian operator group G over E, then xE is a cycle in $D(G;E)$ and every cycle in $D(G;E)$ has this form. The order $n(x)$ of the element x is defined as the order $n(xE)$ of the cycle xE in $D(G;E)$. The order of the element x in G does not exceed n if, and only if, $xP^n = 0$. Mapping the element e in E upon the element xe in xE constitutes an isomorphism of $E/P^{n(x)}$ upon xE which maps ideals upon admissible subgroups. If Z is a cycle in $D(G;E)$, and if $0 \leqq i \leqq n(Z)$, then ZP^i is the uniquely determined subcycle of order $n(Z) - i$ of Z.[6]

Every admissible subgroup of the primary abelian operator group G is a direct sum of cycles (= is the sum of an independent set of cycles = possesses a basis).[6a] If U and V are subgroups in $D(G;E)$, and if $U \leqq V$, then the following condition is necessary and sufficient for U to be a direct summand of G: if the subcycle $Z \neq 0$ of U is contained in a subcycle of order n of V, then there exists a cycle of order n between Z and U.[7]

1F. *Criterion for Primarity.*[8] Suppose that the ring E contains an identity element 1, that E is a cycle[9] of order m in the partially ordered set of all the right-ideals in E, that the abelian group G admits the elements in E as operators, and that[10] $D(G;E)$ contains at least two independent cycles of order m. Then E is primary if, and only if, the following property is satisfied by $D(G;E)$:

If U and V are E-admissible subgroups of G, if $U \leqq V$, and if there exist at most two different subcycles of order 1 of V/U (in $D(G/U;E)$), then (and only then) is V/U a cycle in $D(G/U;E)$.

1G. *Projectivities* are biunivoque and monotone increasing maps of one partially ordered set upon another partially ordered set. If G is a primary abelian

[5] Baer (4), (i) to (v) of Section II.2, p. 304.

[6] Baer (r), Theorem II.2.1.

[6a] Baer (4), Theorem A. of Section II.2, p. 310.

[7] The most important special case of this theorem is the fact that every subcycle of maximum order is a direct summand; cp. Baer (4), Theorem I.3.6, II.2.4 and the remarks on p. 310.

[8] Baer (4), Theorem II.2.4.

[9] This hypothesis is readily seen to imply that every right-ideal in E is a two-sided ideal.

[10] This hypothesis is only needed to show the sufficiency of the condition.

operator group over E, if $D(G;E)$ contains at least three independent cycles of order $m(E)$, and if φ is a projectivity of $D(G;E)$ upon $D(G';E')$ for G'an abelian group admitting the elements in the ring E' as operators, then there exists[11] an isomorphism η of G upon G' and an isomorphism η of E upon E' such that $S^\eta = S^\varphi$ for S in $D(G;E)$ and $(ge)^\eta = g^\eta e^\eta$ for g in G and e in E.

2. The Annulets

Suppose that the ring E contains an identity element 1, and that the abelian group G admits the elements in E as operators. Then the transformation φ shall be termed an *automorphism of* G (more accurately: *an E-automorphism*), if it is a single-valued function of the elements in G with values in G satisfying $(u \pm v)^\varphi = u^\varphi \pm v^\varphi$, $(ue)^\varphi = u^\varphi e$ for u, v in G and e in E. Defining addition and multiplication of automorphisms in the usual fashion it is readily seen that the automorphisms of G form a ring:[12] *the automorphism ring* $\Lambda = \Lambda(G) = \Lambda(G, E)$ *of the group G over E.*

It is the main object of this section to exhibit a class of left-ideals and a class of right-ideals in Λ which each reflect faithfully all the properties of the set $D(G;E)$ of the E-admissible subgroups of G.

If Σ is any subset of Λ, then the *left-annulet* $\Lambda(\Sigma)$ determined by Σ consists of all the automorphisms ξ in Λ satisfying $\xi\Sigma = 0$; and the *right-annulet* $\mathrm{P}(\Sigma)$ consists of all the automorphisms χ in Λ such that $\Sigma\chi = 0$. If T is a subset of G, then the *left-annulet* $\Lambda(T)$determined by T consists of all the automorphisms of G which map G into part of T; and the *right-annulet* $\mathrm{P}(T)$ determined by T consists of all those automorphisms of G which map T upon 0.[13] We observe that left-annulets are left-ideals in Λ and that right-annulets are right-ideals in Λ. The left- and right-annulet determined by the same E-admissible subgroup T of G are connected by the following fundamental relation:[14]

$$\Lambda(T)\mathrm{P}(T) = 0.$$

If $N(\Sigma)$ is the set of all the elements in G which are mapped upon 0 by all the automorphisms in the subset Σ of Λ, then it is readily seen that $N(\Sigma)$ is an E-admissible subgroup of G; the set G^Σ, however, of all the elements g^σ for g in G and σ in Σ need not be a subgroup.

THEOREM 2.1[15]: $\mathrm{P}(\Sigma) = \mathrm{P}(G^\Sigma)$ *and* $\Lambda(\Sigma) = \Lambda(N(\Sigma))$ *for every subset* Σ *of* Λ.

PROOF: The automorphism ξ of G belongs to $\mathrm{P}(\Sigma)$ if, and only if, $\Sigma\xi = 0$; and this is the case if, and only if, $0 = G^{\Sigma\xi} = (G^\Sigma)^\xi$, i.e. if, and only if, ξ is in $\mathrm{P}(G^\Sigma)$; and this proves the first equality. Likewise ξ belongs to $\Lambda(\Sigma)$ if, and

[11] Baer (4), Theorem II.3.1.

[12] Cp. Baer (1), Fitting (1), Shoda (1).

[13] The definitions of $\mathrm{P}(T)$ and $\Lambda(T)$ are due to Shoda (1).

[14] The fundamental importance of this relation has been discovered by Shiffman (1) who based his discussion of the fully characteristic subgroups of primary abelian groups on the consideration of pairs of two-sided ideals defined by this relation.

[15] This is the generalization of a theorem proved by Shiffman (1) for two-sided ideals Σ.

only if, $\xi\Sigma = 0$; or equivalent: if, and only if, $G^{\xi\Sigma} = 0$; and this happens if, and only if, $G\xi \leqq N(\Sigma)$ showing the validity of the second equality.

THEOREM 2.2:[16] *If G is a primary abelian operator group over the (primary) ring E, and if S is an E-admissible subgroup of G, then*

$$G^{\mathrm{A}(S)} = S = N(\mathrm{P}(S)) \text{ and} \tag{1}$$

$$\mathrm{A}(S) = \mathrm{A}(\mathrm{P}(S)), \qquad \mathrm{P}(S) = \mathrm{P}(\mathrm{A}(S)). \tag{2}$$

PROOF: It is an immediate consequence of the definition of $\mathrm{A}(S)$ that $G^{\mathrm{A}(S)} \leqq S$. To prove the desired equality we consider an element v of maximum order in G. There exists by **1E** an E-admissible subgroup V of G such that G is the direct sum of V and of the cycle vE. If s is any element in S, then there exists one and only one automorphism φ of G which maps V upon 0 and v upon s, since $n(s) \leqq n(v)$, and since sE and $E/P^{n(s)}$, vE and $E/P^{n(v)}$ are isomorphic. Clearly φ belongs to $\mathrm{A}(S)$. Hence s is an element in $G^{\mathrm{A}(S)}$ so that $S = G^{\mathrm{A}(S)}$.

It follows from the definition of $\mathrm{P}(S)$ that S is mapped upon 0 by $\mathrm{P}(S)$ and that therefore $S \leqq N(\mathrm{P}(S))$. To prove the desired equality let t be any element in G not in S, and W a greatest E-admissible subgroup of G which contains S, but not t. Every E-admissible subgroup of G which contains W as a proper subgroup contains therefore the element t so that the element $W + t$ generates the only subcycle of order 1 of G/W. Applying the fundamental property mentioned under **1F** we deduce that G/W is a cycle whose order certainly does not exceed the maximum order of the elements in G. Thus there exists in G an element h of order $n(G/W)$; and there exists one and only one E-automorphism γ of G which maps W upon 0 and some generating coset of G/W upon h. Clearly γ belongs to $\mathrm{P}(S)$, since $S \leqq W$; and t does not belong to $N(\mathrm{P}(S))$, since t^{γ} is, exactly as $W + t$, of order 1. This completes the proof of (1).

From (1) and Theorem 2.1 we deduce that $\mathrm{P}(\mathrm{A}(S)) = \mathrm{P}(G^{\mathrm{A}(S)}) = \mathrm{P}(S)$ and that $\mathrm{A}(\mathrm{P}(S)) = \mathrm{A}(N(\mathrm{P}(S))) = \mathrm{A}(S)$.

It is an immediate consequence of Theorem 2.1 and Theorem 2.2, that we need not distinguish between the annulets determined by subsets of G and the annulets which originate from subsets of A.

If we define as *dualities* the biunivoque and monotone decreasing maps of one partially ordered set upon another partially ordered set, then we may restate the results obtained in this section in the form of the following *fundamental theorem*:

If G is a primary abelian operator group over the (primary) ring E, then a projectivity of $D(G;E)$ upon the partially ordered set of all the left-annulets in $\mathrm{A}(G;E)$ is defined by mapping the E-admissible subgroups S of G upon the left-annulet $\mathrm{A}(S)$; and a duality of $D(G;E)$ upon the partially ordered set of all the right-annulets in $\mathrm{A}(G;E)$ is defined by mapping the E-admissible subgroup S of G upon the right-annulet $\mathrm{P}(S)$.

REMARK: If G is an abelian group which admits as operators the elements in

[16] This is the generalization of a theorem proved by Shiffman (1) for the special case of fully characteristic subgroups S.

the ring E, if the ring E meets all the requirements for primary rings (**1D**) with the exception of the postulate that a power of P be 0 though every element in G is annihilated by some power of P, then at least one of the following three conditions is satisfied by the E-admissible subgroups of G:
(A) The orders of the cycles in $D(G;E)$ are bounded.
(B) There exists an E-admissible direct summand of G which contains one and only one subcycle of order n, for every positive integer n.
(C) To every positive integer n there exists a direct summand of G which is a cycle in $D(G;E)$ whose order exceeds n.
If (A) is satisfied by $D(G;E)$, then all the results of this section are obviously still valid. If (B) is satisfied, then one may show by an argument similar to the one we used in the proof of Theorem **2.2** that $S = N(\mathrm{P}(S))$ for every S in $D(G;E)$, though S need not equal $G^{\Lambda(S)}$; and if (C) is satisfied by $D(G;E)$, then $S = G^{\Lambda(S)}$, though S and $N(\mathrm{P}(S))$ may be different.

Chapter II: The Automorphisms of the Automorphism Ring

3. Projectivities of the Group and Automorphisms of the Ring

The main object of this section is the proof of the theorem[17] that

The group of projectivities of the partially ordered set $D(G;E)$ of all the E-admissible subgroups of the primary abelian operator group G over the primary ring E and the group of all the automorphisms of the ring $\Lambda(G;E)$ are essentially the same, provided there exist at least three independent cycles of order[18] *$m(E)$ in $D(G;E)$.*

In order to construct the isomorphism linking these two groups to each other we have to consider the *D-automorphisms of G*: these are biunivoque correspondences between the elements in G which preserve addition and which map E-admissible subgroups upon E-admissible subgroups; the D-automorphisms are therefore proper automorphisms of G which induce projectivities of $D(G;E)$ upon itself. If $D(G;E)$ contains at least two independent cycles of order $m(E)$, then a necessary and sufficient condition[19] for the proper automorphism φ of G to be a D-automorphism is the existence of an automorphism φ of the ring E such that $(ge)^{\varphi} = g^{\varphi}e^{\varphi}$ for g in G and e in E. It is readily deduced from this criterion that $\varphi^{-1}\gamma\varphi$ is an E-automorphism of G, whenever φ is a D-automorphism and γ an E-automorphism of G. This shows that each D-automorphism φ induces a projectivity: $S \rightarrow S^{\Phi}$ of $D(G;E)$ and an automorphism $\gamma \rightarrow \varphi^{-1}\gamma\varphi$ of $\Lambda(G;E)$, if there exist at least two independent cycles of order $m(E)$ in $D(G;E)$ and we shall have attained our object as soon as we have shown that the correlation thus established between projectivities of $D(G;E)$ and automorphisms of $\Lambda(G;E)$ is actually an isomorphism between the two groups, provided there exist at least three independent cycles of order $m(E)$ in $D(G;E)$.

If G is an abelian group, then an *endomorphism of G* is a single-valued function

---

[17] The implications of this theorem as well as some of the relations to known theorems are discussed in section **4**.

[18] For the definition of m(E) cf. **1D**.

[19] This may be derived easily from Baer (4), Theorem II.2.2. or Theorem II.3.1.

of the elements in G which maps elements in G upon elements in G and which preserves addition.

THEOREM 3.1: *If G is a primary abelian operator group over the (primary) ring E, if $D(G;E)$ contains at least two independent*[20] *cycles of order $m(E)$, and if φ is an endomorphism of G, then each of the following properties of φ implies all the others:*

(1) *There exists an element e in E such that $g^\varphi = ge$ for every g.*

(2) *$S^\varphi \leqq S$ for every E-admissible subgroup S of G.*

(3) *$\varphi\alpha = \alpha\varphi$ for every E-automorphism α of G.*

PROOF: It is obvious that (2) and (3) are both consequences of (1). Suppose now that (3) is satisfied by φ and that the E-admissible subgroup S of G is a direct summand of G. Then there exists an E-automorphism γ of G which maps G upon S and which leaves invariant every element in S. If s is any element in S, then $s^\varphi = s^{\gamma\varphi} = s^{\varphi\gamma}$ is an element in S so that both (3) and (2) imply the following condition:

(2′) *$S^\varphi \leqq S$ for every E-admissible direct summand S of G.*

Suppose now that (2′) is satisfied by φ. If b is an element of maximum order in G, then it follows from **1E** that the cycle bE of order $m(E)$ is a direct summand of G; and hence there exists one (and only one) element $e(b)$ in E such that $b^\varphi = be(b)$. If b' and b'' are two independent elements of maximum order, then $b' + b''$ is an element of maximum order; and it follows from $b'e(b') + b''e(b'') = b'^\varphi + b''^\varphi = (b' + b'')^\varphi = (b' + b'')e(b' + b'')$ that $e(b') = e(b' + b'') = e(b'')$. By hypothesis there exist two independent elements v and w of maximum order in G. If b is any element of maximum order in G, then b is independent of at least one[21] of the elements v and w; and we deduce from the fact just derived that $e(b) = e(v) = e(w) = e$ for every element b of maximum order in G. If $g \neq 0$ is any element in G, then g is independent[22] of at least one of the elements v and w. If g and v are independent, then $v + g$ is an element of maximum order so that $ve + g^\varphi = v^\varphi + g^\varphi = (v + g)^\varphi = (v + g)e = ve + ge$ or $g^\varphi = ge$ for every g. Thus (1) is a consequence of (2′), completing the proof of the equivalence of (1), (2) and (3).

COROLLARY 3.2: *Suppose that G is a primary abelian operator group over the (primary) ring E and that $D(G;E)$ contains at least two independent cycles of order $m(E)$.*

(a) *Two D-automorphisms of G induce the same projectivity in $D(G;E)$ if, and only if, they induce the same automorphism in the ring $\Lambda(G;E)$.*

(b) *Mapping the projectivity induced by the D-automorphism φ in $D(G; E)$ upon the automorphism induced by φ in $\Lambda(G;E)$ constitutes an isomorphism of a group*[23] *of projectivities of D upon a group*[23] *of automorphisms of Λ.*

---

[20] It is readily seen that this hypothesis cannot be omitted.

[21] Since cycles different from 0 contain one and only one cycle of order 1, i.e. contain one and only one smallest admissible subgroup different from 0.

[22] Since gE is a cycle in $D(G;E)$.

[23] Which may or may not be the whole group.

The statement (a) is easily deduced from the equivalence of the properties (2) and (3) of Theorem 3.1; and the statement (b) is an immediate consequence of (a).

LEMMA 3.3: *If G is a primary abelian operator group over the (primary) ring E, if $D(G;E)$ contains two independent cycles of order $m(E)$, and if the automorphism φ of the ring $\Lambda(G;E)$ leaves all the left-annulets in Λ invariant, then $\varphi = 1$.*

PROOF: It is an immediate consequence of Theorem 2.2 that φ leaves all the right-annulets in Λ invariant; and that the E-automorphism α of G maps therefore the element g upon 0 if, and only if, α^φ maps g upon 0; and α maps G upon the E-admissible subgroup S of G if, and only if, α^φ maps G upon S.

There exists, by **1E**, a basis B of G. To every element b in B and to every element g in G with $n(g) \leqq n(b)$ there exists one and only one E-automorphism $\alpha(b,g)$ of G which maps b upon g and which maps upon 0 all the other elements in B. Then it follows from the remarks in the preceding paragraph of the proof that $\alpha(b, g)^\varphi$ maps the element b upon a generator of gE hence upon an element $ge(b, g)$ for $e(b, g)$ in E, but not in[24] P, and maps all the other elements in B upon 0.

From $\alpha(b, b) = \alpha(b, b)^2$ we infer $\alpha(b, b)^\varphi = \alpha(b, b)^\varphi$ and, by **1.D**, $e(b, b) \equiv e(b, b)^2$ or $e(b, b) \equiv 1 \bmod P^{n(b)}$, i.e. $\alpha(b, b) = \alpha(b, b)^\varphi$.

If b and g are independent elements, then we infer from $\alpha(b, b + g) = \alpha(b, b) + \alpha(b, g)$ that $be(b, b + g) + ge(b, b + g) = (b + g)e(b, b + g) = b^{\alpha(b,b+g)^\varphi} = b^{\alpha(b,b)^\varphi+\alpha(b,g)^\varphi} = b + ge(b, g)$ and that therefore $e(b, b + g) \equiv 1 \bmod P^{n(b)}$ and $e(b, b + g) \equiv e(b, g) \bmod P^{n(g)}$. Hence $e(b, g) \equiv 1 \bmod P^{n(g)}$ and $\alpha(b, g)^\varphi = \alpha(b, g)$.

To every element b in B there exists an element w in G which is independent of b and has the same order as b. If r is any element in E, then b and $g = br + w$ as well as b and w are independent. Hence we may deduce from the result of the preceding paragraph that $\alpha(b, br)^\varphi = \alpha(b, br)^\varphi + \alpha(b, w) - \alpha(b, w) = \alpha(b, br)^\varphi + \alpha(b, w)^\varphi - \alpha(b, w) = \alpha(b, br + w)^\varphi - \alpha(b, w) = \alpha(b, br + w) - \alpha(b, w) = \alpha(b, br)$.

To every element g in G whose order does not exceed $n(b)$ there exists an element r in E and an element w in G whose order does not exceed $n(b)$ and which is either 0 or independent of b such that $g = br + w$. Hence $\alpha(b, g)^\varphi = \alpha(b, br + w)^\varphi = \alpha(b, br)^\varphi + \alpha(b, w)^\varphi = \alpha(b, br) + \alpha(b, w) = \alpha(b, g)$.

If δ is any E-automorphism of G, then $\delta - \alpha(b, b^\delta)$ maps the basis element b upon 0. It follows therefore from a result stated in the first paragraph of this proof that $\delta^\varphi - \alpha(b, b^\delta) = \delta^\varphi - \alpha(b, b^\delta)^\varphi = (\delta - \alpha(b, b^\delta))^\varphi$ maps b upon 0; and this shows that both δ and δ^φ map the basis element b upon b^φ; and this implies the equality of δ and δ^φ, since they are both E-automorphisms of G. This completes the proof of the fact that $\varphi = 1$.

THEOREM 3.4: *If G_i is, for $i = 1, 2$, a primary abelian operator group over the*

[24] For the definition of the ideal P in E, cp. **1D**.

(primary) ring E_i, if $D_i = D(G_i\,;E_i)$ contains at least three independent cycles of order $m(E_i)$, then there exists to every isomorphism φ of the ring $\mathrm{A}_1 = \mathrm{A}(G_1\,;E_1)$ upon the ring $\mathrm{A}_2 = \mathrm{A}(G_2\,;E_2)$ an isomorphism η of G_1 upon G_2 and an isomorphism η of E_1 upon E_2 such that $(ge)^\eta = g^\eta e^\eta$ for g in G_1 and e in E_1 and such that $\alpha^\varphi = \eta^{-1}\alpha^\eta$ for every α in A_1.

REMARK: A special case of this theorem is the fact that a ring A is the ring of E-automorphisms of essentially at most one operator group G meeting our requirements.[25]

PROOF: It is a consequence of Theorem 2.2 that mapping the left-annulet Λ in A_1 upon Λ^φ constitutes a projectivity of the partially ordered set of all the left-annulets in A_1 upon the partially ordered set of all the left-annulets in A_2. If S is an E_1-admissible subgroup of G_1, then $S^{\varphi*} = G_2^{\mathrm{A}(S)\varphi}$ is by Theorem 2.2, an E_2-admissible subgroup of G_2; and it follows from the theorems of section **2** that φ^* is a projectivity of D_1 upon D_2. There exists by the theorem[26] in **1G** an isomorphism η of G_1 upon G_2 and an isomorphism η of E_1 upon E_2 such that $(ge)^\eta = g^\eta e^\eta$ for g in G_1, e in E_1 and $S^\eta = S^{\varphi*}$ for S in D_1; and we deduce from Theorem 2.2 that

$$\eta^{-1}\mathrm{A}(S)\eta = \mathrm{A}(S^\eta) = \mathrm{A}(S^{\varphi*}) = \mathrm{A}(G_2^{\mathrm{A}(S)\varphi}) = \mathrm{A}(S)^\varphi \quad \text{for } S \text{ in } D_1 .$$

Applying Lemma 3.3 we find now that $\alpha^\varphi = \eta^{-1}\alpha\eta$ for α in A_1, since the two isomorphisms under consideration effect the same projectivity in the partially ordered set of the left-annulets in A_1.

COROLLARY 3.5: *If G is a primary abelian operator group over the primary ring E, and if $D(G;E)$ contains at least three independent cycles of order $m(E)$, then every automorphism of $\mathrm{A}(G;E)$ and every projectivity of $D(G;E)$ upon itself is induced by a D-automorphism of G.*

This is an immediate consequence of Theorem 3.4 and of the Theorem in **1G**. The existence of the desired isomorphism of the group of projectivities of D and the group of automorphisms of A is now assured by Corollaries 3.2 and 3.5. This completes the proof of the theorem enunciated at the beginning of this section.

Since the left-ideals of the ring H_n of all the n by n matrices with integral coefficients form a projection of the system of subgroups of a direct sum of n infinite cyclic groups, and since every projection of this latter system upon itself may be induced by an isomorphism of the underlying group,[27] one may prove by a method similar to, though much cruder than the one used in our proof of Theorem 3.4 that *every automorphism of the ring H_n is an inner automorphism.*

---

[25] Cp. Asano (1), Satz 12, p. 245.

[26] It has been pointed out (Baer (2)) that this theorem would not be true without the hypothesis of the existence of three independent cycles of maximum order. In the proof of Lemma 3.8 we made use of the existence of at least two independent cycles of maximum order. Whether or not either of these hypotheses is needed in order that every automorphism of $\mathrm{A}(G;E)$ is induced by a $D(G;E)$-automorphism of G, is still an open question.

[27] Baer (2), Theorem 13.1, p. 39.

4. Endomorphisms of the Group and Automorphisms of the Ring

If G is an abelian group, then we denote by $\mathrm{H} = \mathrm{H}(G)$ *the ring of all the endomorphisms of* G. If the group G admits the elements in the ring E as operators, then the ring $\mathrm{A}(G;E)$ of the E-automorphisms of G is a subring of $\mathrm{H}(G)$; and if e is any element in E, then an endomorphism of G is defined by mapping the element g in G upon ge. If the ring E is primary, and if $D(G;E)$ contains cycles of order $m(E)$, then $Ge = 0$ implies $e = 0$, since $te = 0$ implies, by **1E**, $e = 0$ for t an element of order $m(E)$; and hence we may identify the elements in E with the endomorphisms they induce in G so that E may be considered another subring of H. If E is a primary ring, and if $D(G;E)$ contains at least two independent cycles of order $m(E)$, then it follows from Theorem 3.1 that the subrings E and A of H are each the centralizer of the other one in H.[28] This implies in particular that an element in H which possesses an inverse in H transforms E into itself if, and only if, it transforms A into itself; and this fact may be stated as follows: *the subrings* A *and* E *of* H *have the same normalizing group* N *in* H.

THEOREM 4.1: *If G is a primary abelian operator group over the (primary) ring E, and if $D(G;E)$ contains at least three independent cycles of order $m(E)$, then the normalizing group* N *of E and* A *in* H *consists exactly of the D-automorphisms of G and all the automorphisms of E and of* A *are induced by elements in* N.

PROOF: It has been pointed out at the beginning of section 3, that the D-automorphisms of G transform the ring A into itself, i.e. they belong to N. If conversely ϑ belongs to N, then there exists an inverse endomorphism to the endomorphism ϑ so that ϑ is a proper automorphism of G; and it is a consequence of Corollary 3.5 that there exists a D-automorphism ρ of G satisfying: $\rho^{-1}\alpha\rho = \vartheta^{-1}\alpha\vartheta$ for every α in A. Hence it follows from Theorem 3.1 that $\vartheta\rho^{-1}$ belongs to E and induces the identical projectivity in $D(G;E)$. Thus ρ and $\vartheta\rho^{-1}$ are D-automorphisms of G and consequently every ϑ in N is a D-automorphism. It is a consequence of Corollary 3.5 that every automorphism of A is induced by D-automorphisms of G; and it is well known[29] that every automorphism of E is induced by a D-automorphism of G. Hence every automorphism of the rings E and A is induced by elements in N.

LEMMA 4.2: *Suppose that U and V are subrings of the ring W, that U is the centralizer of V and V the centralizer of U in W, and that the element r in W possesses an inverse in W. Then r induces in V an inner automorphism of V if, and only if, it induces in U an inner automorphism of U.*

PROOF: If r induces in U an inner automorphism of U, then there exists an element s in U which possesses an inverse in U and which satisfies $s^{-1}us = r^{-1}ur$ for every u in U. The element $t = rs^{-1}$ belongs to the centralizer of U so that t is an element in V which possesses an inverse in V. Since s belongs to the centralizer U of V, we find that $t^{-1}vt = r^{-1}vr$ for every v in V, i.e. r induces in V an inner automorphism of V.

---

[28] This implies in particular the equality of the centrum of $\mathrm{A}(G;E)$ and of the centrum of the ring E, a fact that has already been noted by K. Asano (1), Satz 7, p. 242.

[29] This is an immediate consequence of the existence of a basis of G; cp. **1E**.

THEOREM 4.3: *Suppose that G is a primary abelian operator group over the (primary) ring E, and that $D(G;E)$ contains at least three independent cycles of order $m(E)$. Then every automorphism of the ring $\Lambda(G;E)$ is an inner automorphism of Λ if, and only if, every automorphism of the ring E is an inner automorphism of E.*

This is an immediate consequence of Theorem 4.1 and Lemma 4.3, and of the fact that E and Λ are each the centralizer of the other in H.

We mention two interesting special cases:

1. If G is an ordinary primary abelian group the orders of whose elements are bounded and which contains at least three independent elements of maximum order, then every automorphism of the ring Λ of automorphisms of G is an inner automorphism of Λ.

For E is the ring of integers modulo a power of a prime.

2. If n is a finite or infinite cardinal number not less than 3, and if Λ is the ring of all the row-finite, n by n matrices with coefficients from the field of real numbers (or the field of quaternions or some other field all of whose automorphisms are inner), then every automorphism of Λ is an inner automorphism.

For we may choose for G the group of rank n over the field of coefficients.

Every inner automorphism of a ring is a center automorphism[30] of this ring; and it is a well known theorem[31] that all center automorphisms of a finite simple algebra are inner automorphisms. Finite simple algebras are known[32] to be rings of square matrices with coefficients from a field which is finite over its center; and thus they are automorphism rings of primary abelian operator groups. The following example shows that automorphism rings of primary abelian operator groups may possess center automorphisms which are not inner automorphisms in spite of finiteness hypotheses.

It is a consequence of Lemma 4.2 that it suffices to construct a *primary ring E which is finite over its center and which possesses outer center automorphisms.*

Denote by F a commutative field whose characteristic is not 2 and by Q the field $F(x)$ of all the rational functions in an indeterminate x with coefficients in F. There exists one and only one automorphism φ of Q which maps x upon $-x(\neq x)$ and which leaves invariant all the elements in F. If q is an element not 0 in Q, then $q^{\varphi}q^{-1}$ is readily verified to be an element different from x and x^{-1}.

Denote by E the set of all the ordered pairs (a, b) of elements in Q subject to the following rules: $(a, b) = (c, d)$ if, and only if, $a = c$ and $b = d$; $(a, b) \pm (c, d) = (a \pm c, b \pm d)$, $(a, b)(c, d) = (ac, bc^{\varphi} + ad)$. It is easily seen that E is a ring with identity $1 = (1, 0)$, that the element (a, b) in E possesses an inverse in E if, and only if, $a \neq 0$; and that the only ideal (right- or left-) in E which is different from 0 and E is the two-sided ideal P consisting of all the elements $(0, b)$. Thus E is a primary ring whose center consists exactly of the elements $(a, 0)$ for a in $F(x^2)$, showing that the rank of E over its center is exactly 4. An

[30] i.e. an automorphism which leaves invariant every element in the center of this ring.

[31] Brauer (1), Skolem (1).

[32] Cp. e.g. Albert (1), Theorem 9, p. 39.

automorphism ψ of E is defined by $(a, b)^{\psi} = (a, bx)$ and this automorphism is a center automorphism, since $(a, 0) = (a, 0)^{\psi}$. If (r, s) for $r \neq 0$ were to induce the automorphism ψ in E, then $(0, r) = (r, s)(0, 1) = (0,1)^{\psi}(r, s) = (0, x)(r, s) = (0, xr^{\varphi})$ or $x = r(r^{\varphi})^{-1}$; and the impossibility of this equation has been pointed out before. Thus E meets all the requirements.

CHAPTER III: IDEAL-THEORETICAL PROPERTIES OF THE ANNULETS

5. Finite Sums of Cycles and Finite Cross-cuts of Anti-cycles

An element s in the partially ordered set S is termed an *anti-cycle of index n in S*, if the elements in S which contain s form a cycle of order n. The E-admissible subgroup V of the primary abelian operator group G over the (primary) ring E is, by the Fundamental Theorem of section **2**, a cycle (an anticycle) in $D(G;E)$ if, and only if, $\Lambda(V)$ is a cycle (an anti-cycle) in the partially ordered set of all the left-annulets in $\mathrm{A}(G;E)$; and this is the case if, and only if, $\mathrm{P}(V)$ is an anti-cycle (a cycle) in the partially ordered set of all the right-annulets in $\mathrm{A}(G;E)$.[33]

LEMMA 5.1: *If G is a primary abelian operator group over the (primary) ring E, if Z is a cycle and S an anti-cycle in $D(G;E)$, then every E-automorphism of Z is induced by an E-automorphism of G; and every E-automorphism of G/S is induced by an E-automorphism of G which maps S into part of S.*

PROOF: If $Z = zE$, and if φ is an E-automorphism of Z, then $z^{\varphi} = ze$ for some e in E and φ is completely determined by e. There exists by **1E** a basis B of G and hence there exist elements $d(b)$ in E almost all of which are 0 such that $z = \sum_{b \text{ in } B} bd(b)$. There exist elements $e(b)$ in E satisfying $d(b)e = e(b)\, d(b)$, since E is a primary ring and its left-ideals are right-ideals. There exists finally one and only one E-automorphism γ of G which maps b upon $be(b)$; and γ induces φ in Z, since γ maps z upon ze.

The basis B contains at least one element, say w, which generates G modulo S, since otherwise all the cosets $S + b$ for b in B would be contained in $(G/S)P$ (by **1E**) so that B would be part of $GP < G$, an impossibility. Since the elements in the cycle wE represent all the cosets of G/S, there exists to every element $b(\neq w)$ in B an element $r(b)$ in E such that $b - wr(b)$ is in S. If p is a prime in E (**1D**) and if n is the order of the cycle G/S, then wp^n is an element in S. The E-admissible subgroup S^* generated by wp^n and the elements $b - wr(b)$ is part of S; and G/S^* is a cycle whose order does not exceed n, since every coset in G/S^* is represented by an element in the cycle wE, and since $wP^n \leqq S^*$. Consequently $S = S^*$.

If $W = S + w$, and if σ is an E-automorphism of G/S, then $W^{\sigma} = We$ for e in E; and σ is completely determined by e. There exist elements $c(b)$ in E such that $er(b) = r(b)c(b)$, since E is a primary ring and its right-ideals are left-ideals; and

[33] If S is an anti-cycle in $D(G;E)$, z an element in G, but not in $S + GP$, then $G = S + zE$ where zE is a cycle in $D(G;E)$ and the index of S in $D(G;E)$ does not exceed the order of zE.

there exists one and only one E-automorphism τ of G which maps w upon we and $b \neq w$ in B upon $bc(b)$. Since $(b - wr(b))^\tau = bc(b) - wer(b) = (b - wr(b))c(b)$, it follows that τ maps the elements $b - wr(b)$, wp^n which generate S upon elements in S so that $S^\tau \leqq S$. The E-automorphism τ maps consequently every coset of G/S into part of a well-determined coset of G/S and τ maps in particular the generating coset W into part of We. Thus we have shown that σ is induced by τ in G/S.

LEMMA 5.2: *If G is a primary abelian operator group over the (primary) ring E, and if G is a cycle of order n in $D(G;E)$, then $A(G;E)$ and E/P^n are anti-isomorphic rings.*

PROOF: There exists an element g in G such that $G = gE$. If φ is any E-automorphism of G, then $g^\varphi = ge$ where $P^n + e$ is uniquely determined by g and φ and where g and e determine φ completely. If η is another E-automorphism of G, $g^\eta = gd$, then $g^{\varphi\eta} = (ge)^\eta = g^\eta e = gde$; and now it is obvious how to complete the proof of our statement.

If G is a primary abelian operator group over the (primary) ring E, and if G/S for S an E-admissible subgroup is the sum of a finite number of cycles in $D(G/S;E)$, then it follows from **1E** that G/S is the direct sum of a finite number of cycles and therefore S is the cross-cut of a finite number of anti-cycles. From this fact one deduces easily the equivalence of the following three properties of the E-admissible subgroup S of G: (1) G/S is the sum of a finite number of cycles in $D(G;E)$. (2) S is the cross-cut of a finite number of anti-cycles in $D(G;E)$. (3) $P(S)$ is the sum of a finite number of cycles in the partially ordered set of all the right-annulets in $A(G;E)$.

LEMMA 5.3: *If G is a primary abelian operator group over the (primary) ring E, if the E-admissible subgroup S of G is the sum of a finite number of cycles in $D(G;E)$, and if the E-admissible subgroup T of G is the cross-cut of a finite number of anti-cycles in $D(G;E)$, then S is part of an E-admissible direct summand of G which is the sum of a finite number of cycles; and T contains an E-admissible direct summand of G which is the cross-cut of a finite number of anti-cycles in $D(G;E)$.*

PROOF: Our statement concerning S is an almost obvious consequence of the fact that there exists a basis of G (**1E**). Denote by W the set of all the E-admissible subgroups H of G which are direct summands of G and which are cross-cuts of a finite number of anti-cycles in G. There exists a subgroup K in W such that $(T + K)/T$ is as small as possible, since the minimum condition is satisfied by the E-admissible subgroups of G/T. If K were not part of T, then $T < T + K$ and there would exist an E-admissible subgroup V between T and $T + K$ such that $(T + K)/V$ is a cycle of order 1. Since K is not part of V, there exists in K an element u of least order which is not in V. Clearly $uP \leqq V$ and $V + uE = T + K$. There exists by **1E** a basis B of K and there exists in B a finite number of elements $b_1, \cdots, b_j$ such that u is in $\sum_{i=1}^{j} b_iE$. Denote by B' the set of those elements in B which are different from all the b_i and which are contained in V; and by B'' the set of the elements in B which are

different from the b_i and which are not in V. If b is in B'', then there exists an element $e(b)$ in E such that $b^* = b - ue(b)$ is in V. Since b is in $T + K$, but not in V, it follows that $n(ue(b)) \leqq n(u) \leqq n(b)$. Since u is independent of the set of elements in B' and B'', it follows that a basis B^* of K is formed by the elements b_i, the elements in B' and the elements b^* for b in B''. It is readily verified that the E-admissible subgroup U of G which is generated by the elements in B^* different from b_i belongs to W, though $T + U \leqq V < T + K$, a contradiction which proves our contention.

6. The Finite Automorphisms

If φ is an E-automorphism of the abelian group G over the primary ring E, then G^φ and $G/N(\varphi)$ are isomorphic groups over the operator ring E; thus G^φ is the sum of a finite number of cycles if, and only if, $N(\varphi)$ is the cross-cut of a finite number of anti-cycles. The E-automorphism φ of G is termed *finite*, if G^φ is the sum of a finite number of cycles and $N(\varphi)$ the cross-cut of a finite number of anti-cycles.

LEMMA 6.1: *If G is a primary abelian operator group over the (primary) ring E, and if φ is a finite E-automorphism of G, then there exist E-admissible subgroups U and V of G with the following properties:*

(i) *G is the direct sum of U and V.*

(ii) *U is the sum of a finite number of cycles in $D(G;E)$.*

(iii) *$G^\varphi \leqq U$ and $V \leqq N(\varphi)$.*

PROOF: There exist by Lemma 5.3 E-admissible subgroups A and B of G such that G is the direct sum of A and B, A is the sum of a finite number of cycles in $D(G;E)$ and $B \leqq N(\varphi)$. Denote by C the cross-cut of B and $A + G^\varphi$. Then C is an E-admissible subgroup of B and C is the sum of a finite number of cycles in $D(B;E)$. Hence there exist by Lemma 5.3 E-admissible subgroups V and W of B such that B is the direct sum of V and W, W is the sum of a finite number of cycles in $D(G;E)$ and $C \leqq W$. Then G is the direct sum of $U = A + W$ and V, U is the sum of a finite number of cycles and $G^\varphi \leqq U$, $V \leqq N(\varphi)$.

We denote by $\Phi = \Phi\ (G;E)$ *the set of all the finite E-automorphisms of G.*

LEMMA 6.2: *If G is a primary abelian operator group over the (primary) ring E, then $\Phi\ (G;E)$ is a two-sided, idempotent ideal in the ring $\Lambda(G;E)$; and both the sum of all the left-annulets $\Lambda(Z)$ for cyclic Z and the sum of all the right-annulets $\mathrm{P}(H)$ for anti-cyclic H are equal to $\Phi\ (G;E)$.*

PROOF: If φ and γ are finite E-automorphisms, and if ρ is an E-automorphism of G, then we infer from $G^{\varphi\pm\gamma} \leqq G^\varphi + G^\gamma$ and $G^{\rho\varphi} \leqq G^\varphi$ that Φ is a left-ideal in Λ. Since $G^{\varphi\rho}$ is a linear map of G^φ, both subgroups are generated by a finite number of elements, showing that Φ is a two-sided ideal in Λ. If φ is a finite E-automorphism of G, then there exist by Lemma 6.1 E-admissible subgroups U and V of G meeting the requirements (*i*) to (*iii*)of Lemma 6.1. If ϵ is the E-automorphism of G which maps V upon 0 and leaves invariant every element in U, then ϵ is an idempotent such that $\epsilon\varphi = \varphi\epsilon = \varphi$ and ϵ is finite; and hence Φ is idempotent. It is a consequence of **1E** that $G/N(\varphi)$ is the direct sum of a

finite number of cycles $Z_1, \cdots, Z_k$. There exists one and only one E-automorphism φ_i of G which induces φ in Z_i and which maps the Z_j for $j \neq i$ upon 0. Clearly $\varphi = \varphi_1 + \cdots + \varphi_k$, G^{φ_i} is a cycle and $N(\varphi_i)$ an anti-cycle; and this completes the proof.

We note several consequences of this Lemma 6.2:

1. The identity-automorphism of G is finite if, and only if, G itself is the sum of a finite number of cycles in $D(G;E)$; and Φ is a (right- or left-) annulet if, and only if, $\Phi = \Lambda$.
2. Every right- (left-) ideal in the subring Φ of Λ is a right- (left-) ideal in the ring Λ.
3. If S and T are E-admissible subgroups of G, then each of the following equalities may be easily seen to imply the others:

$S = T$, $\mathrm{P}(S) \cap \Phi = \mathrm{P}(T) \cap \Phi$, $\Lambda(S) \cap \Phi = \Lambda(T) \cap \Phi$.

Comparing 2. and 3. we note that the ideal theory in the rings Φ and Λ is different, provided these rings are different; but that the theory of annulets is essentially the same in both rings.

It is known[34] that the E-automorphisms of G may be approximated by means of finite E-automorphisms. It is the object of the following remarks to describe this approximation in terms which involve properties of the ring Λ only. With this in mind we denote by $\Omega = \Omega(G;E)$ *the set of all the right-annulets* I *in* Λ *which are cross-cuts of a finite number of anti-cycles in the partially ordered set of the right-annulets in* Λ; and it will be convenient to term *Ω-function* any single-valued function $\varphi(\mathrm{I})$ of the right-annulets in Ω with the following properties:

$\varphi(\mathrm{I})$ *is a coset of* Λ/I; *and* $\mathrm{I} \leqq \mathrm{I}'$ *implies* $\varphi(\mathrm{I}) \leqq \varphi(\mathrm{I}')$.

If φ is any E-automorphism of G, then $\mathrm{I} + \varphi$ for I in Ω is the typical example of such an Ω-function.

THEOREM 6.3: *Suppose that G is a primary abelian operator group over the (primary) ring E.*

(a) *The right-annulet* I *is in* Ω *if, and only if,* $\Lambda(\mathrm{I}) \leqq \Phi$.

(b) *To every* I *in* Λ *and* φ *in* Λ *there exists a finite E-automorphism ρ of G such that* $\varphi \equiv \rho \bmod \mathrm{I}$.

(c) *To every Ω-function $\varphi(\mathrm{I})$ there exists one and only one E-automorphism φ of G such that* $\varphi(\mathrm{I}) = \mathrm{I} + \varphi$ *for every* I *in* Ω.

It seems desirable to find an expression of the closure property (c) of the ring Λ which does not involve the elements of Λ, but which refers to the annulets only.

PROOF: If S is an E-admissible subgroup of G, B a basis of S and C a basis of G, then the (cardinal) number of the elements in B whose order is at least n does not exceed the (cardinal) number of the elements in C whose order is at least n. Hence there exists an E-automorphism φ of G such that $G^{\varphi} = S$; and consequently $\Lambda(\mathrm{I}) \leqq \Phi$ if, and only if, $G^{\Lambda(\mathrm{I})}$ is the sum of a finite number of cycles in $D(G;E)$. It is a consequence of Theorems 2.1 and 2.2 that $G^{\Lambda(\mathrm{I})} = N(\mathrm{I})$; and

[34] Baer (1).

that $N(\mathrm{I})$ is the sum of a finite number of cycles in $D(G;E)$ if, and only if, $\mathrm{I} = \mathrm{P}(N(\mathrm{I}))$ is the cross-cut of a finite number of anti-cycles in the partially ordered set of the right-annulets in Λ. It is a consequence of Lemma 5.3 that G is the direct sum of E-admissible subgroups U and V such that $N(\mathrm{I}) \leqq U$ and such that U is the sum of a finite number of cycles in $D(G;E)$. Then there exists to every E-automorphism φ of G one and only one E-automorphism γ of G which coincides with φ on U and which maps V upon 0. It is clear that γ is a finite E-automorphism satisfying $\varphi \equiv \gamma \bmod \mathrm{I}$.

We note that $\mathrm{P}(x)$ is, for every x in G, a right-annulet in the class Ω, since $\Lambda(\mathrm{P}(x)) = \Lambda(xE) \leqq \Phi$. If φ and γ are any two E-automorphisms of G satisfying $\mathrm{I} + \varphi = \mathrm{I} + \gamma$ for every I in Ω, then $\varphi - \gamma = 0$ as an element in the cross-cut 0 of all the $\mathrm{P}(x)$ for x in G. If finally $\varphi(\mathrm{I})$ is an Ω-function, then all the elements in the coset $\varphi(\mathrm{P}(x))$ have the same value x^φ on the element x in G. If x is in G and e in E, then $xeE \leqq xE$ implies $\mathrm{P}(x) = \mathrm{P}(xE) \leqq \mathrm{P}(xeE) = \mathrm{P}(xe)$ so that $(xe)^\varphi = (xe)^{\varphi(\mathrm{P}(x))} = x^{\varphi(\mathrm{P}(x))}\, e = x^\varphi e$. If x and y are elements in G, then $\varphi(\mathrm{P}(xE + yE))$ is part of $\varphi(\mathrm{P}(x))$, $\varphi(\mathrm{P}(y))$ and $\varphi(\mathrm{P}(x \pm y))$ so that $(x \pm y)^\varphi = (x \pm y)^{\varphi(\mathrm{P}(xE+yE))} = x^{\varphi(\mathrm{P}(x))} + y^{\varphi(\mathrm{P}(y))} = x^\varphi \pm y^\varphi$. Thus φ is an E-automophism of G. If I is in Ω and x in $N(\mathrm{I})$, then $\mathrm{I} = \mathrm{P}(N(\mathrm{I})) \leqq \mathrm{P}(x)$ and hence $\varphi(\mathrm{I}) \leqq \varphi(\mathrm{P}(x))$; and this implies that $x^\varphi = x^{\varphi(\mathrm{I})}$ for x in $N(\mathrm{I})$ or $\varphi(\mathrm{I}) = \mathrm{I} + \varphi$.

7. Sums of Annulets

It is easily seen that the cross-cut of any set of right-(left-)annulets is a right-(left-)annulet. Thus there exists to every set of right-(left-)annulets a smallest right-(left-) annulet, the join of the annulets in the set. But in general this join will be different from the sum of the annulets in the set.

THEOREM 7.1: *Suppose that G is a primary abelian operator group over the (primary) ring E, and that P_i is a right-annulet, Λ_i a left-annulet in $\Lambda(D;E)$.*

(a) *If $\Lambda_1 \cap \Lambda_2 = 0$, then $\Lambda_1 + \Lambda_2$ is a left-annulet.*

(b) *If $\mathrm{P}_1 \cap \mathrm{P}_2 = 0$, then $\mathrm{P}_1 + \mathrm{P}_2$ is a right-annulet.*

(c) *If $\mathrm{P}_1 = \mathrm{P}(\Lambda_i)$, $\Lambda_i = \Lambda(\mathrm{P}_i)$ and $\mathrm{P}_1 \cap \mathrm{P}_2 = 0 = \Lambda_1 \cap \Lambda_2$, then $\mathrm{P}_1 + \mathrm{P}_2 = \Lambda = \Lambda_1 + \Lambda_2$.*

PROOF: It is an immediate consequence of the Fundamental Theorem of section **2** that the cross-cut of $S_1 = G^{\Lambda_1}$ and $S_2 = G^{\Lambda_2}$ is $G^{\Lambda_1 \cap \Lambda_2} = 0$, and that the smallest left-annulet which contains both Λ_1 and Λ_2 is $\Lambda(S_1 + S_2)$. If φ is any element in $\Lambda(S_1 + S_2)$ and g any element in G, then g^φ is an element in the direct sum $S_1 + S_2$. Hence there exist uniquely determined elements g_1 and g_2 in S_1 and S_2 respectively such that $g^\varphi = g_1 + g_2$. The transformation φ_i which maps g upon g_i is readily seen to be an E-automorphism in $\Lambda_i = \Lambda(S_i)$; and $\varphi = \varphi_1 + \varphi_2$. Hence $\Lambda_1 + \Lambda_2 = \Lambda(S_1 + S_2)$.

If $T_i = N(\mathrm{P}_i)$, then it is an immediate consequence of the Fundamental Theorem of section **2** that $T_1 + T_2 = N(\mathrm{P}_1 \cap \mathrm{P}_2) = G$ and that $\mathrm{P}(T_1 \cap T_2)$ is the smallest right-annulet which contains both P_1 and P_2. Since $G/(T_1 \cap T_2)$ is the direct sum of $T_1/(T_1 \cap T_2)$ and $T_2/(T_1 \cap T_2)$, there exists to every E-automorphism φ in $\mathrm{P}(T_1 \cap T_2)$ one and only one E-automorphism φ_i which maps T_i upon

0 and which coincides with φ on $T_{i\pm1}/(T_1 \cap T_2)$. Clearly $\varphi = \varphi_1 + \varphi_2$ and φ_i is in $\mathrm{P}_i = \mathrm{P}(T_i)$. Hence $\mathrm{P}_1 + \mathrm{P}_2 = \mathrm{P}(T_i \cap T_2)$.

If the hypotheses of (c) are satisfied by Λ_i, P_i, then it follows from what has already been shown that $S_i = G^{\Lambda_i} = N(\mathrm{P}_i)$, that G is the direct sum of S_1 and S_2, $\Lambda_1 + \Lambda_2 = \Lambda(G) = \Lambda = \mathrm{P}(0) = \mathrm{P}_1 + \mathrm{P}_2$.

If G is an abelian operator group over the ring E, if S is an E-admissible subgroup of G, and if φ is a single-valued additive function of the elements in S with values in G satisfying $(se)^\varphi = s^\varphi e$ for s in S and e in E, then φ is *a linear transformation of S into G.*

LEMMA 7.2: *If the primary abelian operator group G over the (primary) ring E is the direct sum of cycles of equal order m in $D(G;E)$, if S is an E-admissible subgroup of G, and if φ is a linear transformation of S into G, then φ is induced in S by an E-automorphism of G.*

PROOF: If p is a prime in E (cf. **1G**), then every element in G has the form gp^i for g an element of order m in G. Hence there exists a basis B of G and integers $h(b)$ for b in B such that the E-admissible subgroup S of G is generated[35] by the elements $bp^{h(b)}$. Since $n((bp^{h(b)})^\varphi) \leqq n(bp^{h(b)}) = m - h(b)$, there exist elements b' in G such that $b'p^{h(b)} = (bp^{h(b)})^\varphi$. There exists one and only one E-automorphism ψ of G which maps every b in B upon the corresponding b', since $n(b') \leqq n(b) = m$; and ψ clearly induces φ in S.

THEOREM 7.3: *The following three properties of the primary abelian operator group G over the (primary) ring E imply each other:*

(a) *G is the direct sum of cycles of equal order m in $D(G;E)$.*

(b) *The sum of any two right-annulets in $\Lambda(G;E)$ is a right-annulet.*

(c) *The sum of any two left-annulets in $\Lambda(G;E)$ is a left-annulet.*

PROOF: Suppose that condition (a) is satisfied by G. It is a consequence of the Fundamental Theorem in section **2** that any two right annulets in Λ have the form $\mathrm{P}(S)$, $\mathrm{P}(T)$ where S and T are E-admissible subgroups of G, and that $\mathrm{P}(S \cap T)$ is the smallest right-annulet containing both $\mathrm{P}(S)$ and $\mathrm{P}(T)$. If υ is an E-automorphism in $\mathrm{P}(S \cap T)$, then there exists a linear transformation ψ of $S + T$ into G which coincides with υ on S and which maps T upon 0, since $S + T$ is modulo $S \cap T$ the direct sum of S and T, and since $(S \cap T)^\upsilon = 0$. There exists by Lemma 7.2 an E-automorphism ω of G which induces ψ in $S + T$. Thus ω belongs to $\mathrm{P}(T)$ and $\upsilon - \omega$ belongs to $\mathrm{P}(S)$; and hence υ belongs to $\mathrm{P}(S) + \mathrm{P}(T)$; i.e. (a) implies (b).

It is a consequence of the Fundamental Theorem of section **2** that any pair of left-annulets in Λ may be represented in the form $\Lambda(S)$, $\Lambda(T)$ where S and T are E-admissible subgroups of G, and that $\Lambda(S + T)$ is the smallest left-annulet containing both $\Lambda(S)$ and $\Lambda(T)$. If φ is an E-automorphism in $\Lambda(S + T)$, B a basis of G, then b^φ is an element in $S + T$ and there exist therefore elements $s(b)$, $t(b)$ in S and T respectively such that $b^\varphi = s(b) + t(b)$. Since b is by

[35] Note that the elements $bp^{h(b)}$ which are different from 0 form a basis of S. The existence of a basis of S is assured by **1E** and such a basis would be the starting point for the construction of B.

(a) an element of maximum order m in G, there exist uniquely determined E-automorphisms η and κ such that $b^\eta = s(b)$, $b^\kappa = t(b)$. Clearly $\varphi = \eta + \kappa$ and η is in $\Lambda(S)$, κ in $\Lambda(T)$ so that (c) is a consequence of (a).

If condition (a) is not satisfied by the group G, then G is the direct sum of E-admissible subgroups U, V, W where W may be 0 and where U and V are cycles in $D(G; E)$ such that $0 < n(U) < n(V) =$ the maximum-order of the cycles in $D(G; E)$. There exist elements u and v of order 1 in U and V respectively. The cycles uE and $(u + v)E$ are of order 1 and their cross-cut is 0. Thus Λ is the smallest right-annulet containing both $\mathrm{P}(u)$ and $\mathrm{P}(u + v)$. If the identical automorphism 1 were in $\mathrm{P}(u) + \mathrm{P}(u + v)$, then there would exist E-automorphisms η, κ such that $u^\eta = 0$, $(u + v)^\kappa = 0$ and $\eta + \kappa = 1$; and this would imply: $u = u^{\eta+\kappa} = u^\kappa$, $0 = (u + v)^\kappa = u + v^\kappa$ or $uE = (vE)^\kappa \leqq GP^{n(V)-1}$ which is impossible, since $UP^{n(V)-1} \leqq UP^{n(U)} = 0$. Consequently (b) implies (a). If $V = wE$, then $(u + w)E$ is a cycle of order $n(V)$ whose cross-cut with V is exactly VP. Automorphisms in $\Lambda(V)$ and $\Lambda((u + w)E)$ map elements of an order not exceeding $n(V) - 1$ upon elements in VP; and every automorphism in $\Lambda(V) + \Lambda((u + w)E)$ maps therefore the elements of an order not exceeding $n(V) - 1$ upon elements in VP. But there exist E-automorphisms which map U upon uE, W upon 0 and leave invariant every element in V. Such an automorphism would be in $\Lambda(V + (u + w)E) = \Lambda(V + uE)$, but not in $\Lambda(V) + \Lambda((u + w)E)$, since it maps an element in V of an order less than $n(V)$ upon the element u not in VP; and hence (a) is a consequence of (c).

COROLLARY 7.4: *If the primary abelian operator group G over the (primary) ring E is the direct sum of cycles of equal order m in $D(G; E)$, then*

(i) $\varphi\Lambda = \mathrm{P}(N(\varphi))$ *and* $\Lambda\varphi = \Lambda(G^\varphi)$ *for every E-automorphism φ of G;*

(ii) *every right-(left-)annulet is a principal right-(left-)ideal in $\Lambda(G; E)$;*

(iii) *every right-(left-)ideal in Λ which is generated by a finite number of elements*[36] *is a principal right-(left-)ideal in Λ.*

PROOF: If γ is in $\mathrm{P}(N(\varphi))$, then $N(\varphi) \leqq N(\gamma)$. To every element x in G^φ there exists one and only one coset $x^{\varphi^{-1}}$ of $G/N(\varphi)$ which is mapped upon x by φ. Thus $N(\gamma) + x^{\varphi^{-1}}$ is a single, well determined coset of $G/N(\gamma)$ and $x^\eta = (N(\gamma) + x^{\varphi^{-1}})^\gamma$ is a uniquely determined element in G^γ. This transformation η is readily seen to be a linear transformation of G^φ into G; and hence it follows from Lemma 7.2 that η is induced in G^φ by an E-automorphism κ of G. If x is any element in G, then $x^{\varphi\kappa} = x^{\varphi\eta} = (N(\gamma) + x)^\gamma = x^\gamma$ or $\varphi\kappa = \gamma$ and this proves $\varphi\Lambda = \mathrm{P}(N(\varphi))$.

If ρ is an automorphism in $\Lambda(G^\varphi)$, then $G^\rho \leqq G^\varphi$ and the set T of all the elements in G which are mapped by φ upon elements in G^ρ is an E-admissible subgroup such that $N(\varphi) \leqq T$, $T^\varphi = G^\rho$. If Y is any coset of $G/N(\rho)$, then Y^ρ is a well determined element in $G^\rho = T^\varphi$ and there exists one and only one coset $Y^\sigma = Y^{\rho\varphi^{-1}}$ of $T/N(\varphi)$ which is mapped by φ upon the element Y^ρ. It is readily seen that σ is an isomorphism of the group $G/N(\rho)$ over E upon the group $T/N(\varphi)$ over E. There exists a basis B of G. If b is an element in B, then denote by b^*

[36] Has a finite basis in the customary terminology.

some element in $(N(\rho) + b)^{\sigma}$ so that $(N(\rho) + b)^{\sigma} = N(\varphi) + b^*$ for every b in B. Since all the elements b in B are of the maximum-order m in G, there exists one and only one E-automorphism τ of G which maps b upon b^* for every b in B. Then $b^{\tau\varphi} = b^{*\varphi} = (N(\varphi) + b^*)^{\varphi} = (N(\rho) + b)^{\sigma\varphi} = (N(\rho) + b)^{\rho} = b^{\rho}$ or $\tau\varphi = \rho$ and this proves $\Lambda\varphi = \Lambda(G^{\varphi})$.

It is a consequence of the Fundamental Theorem of section **2** that every right-(left-)annulet in Λ has the form $\mathrm{P}(S)$ $(\Lambda(S))$ for S an E-admissible subgroup of G; and it is readily verified that there exist to every E-admissible subgroup S of G E-automorphisms φ, γ such that $S = N(\varphi)$ and $S = G^{\gamma}$. Now (ii) is a consequence of (i). The assertion (iii) finally is an immediate consequence of (i), (ii) and Theorem 7.3.

THEOREM 7.5: *The following five properties of the primary abelian operator group G over the (primary) ring E are equivalent:*

(1) *G is the direct sum of a finite number of cycles of equal order.*[37]

(2) *Sums*[38] *of right-annulets in $\Lambda(G;\ E)$ are right-annulets.*

(3) *Every right-ideal in Λ is a right-annulet.*

(4) *Sums*[38] *of left-annulets in Λ are left-annulets.*

(5) *Every left-ideal in Λ is a left-annulet.*[39]

PROOF: It is obvious that (3) implies (2) and that (5) implies (4). If (2) or (4) is satisfied by G, then G is, by Theorem 7.3, the direct sum of cycles of equal order; and the ideal $\Phi(G;\ E)$ of the finite E-automorphisms of G is by Lemma 6.2 equal to Λ so that in particular the identity is a finite automorphism. But this is the case if, and only if, G is the direct sum of a finite number of cycles, proving that (1) is a consequence of (2) as well as of (4).

If condition (1) is satisfied by the group G, then the maximum and the minimum condition are both satisfied by the E-admissible subgroups of G. Hence it follows from the Fundamental Theorem of section **2** that the maximum condition is satisfied by the right-annulets and by the left-annulets in $\Lambda(G;\ E)$. Furthermore it follows from Corollary 7.4 that annulets are principal ideals and that the sums of a finite number of principal ideals are annulets. Applying the maximum condition for the annulets it is readily deduced that every ideal is a principal ideal and therefore an annulet; and this shows that (3) and (5) are consequences of (1).

8. Products of Annulets

If Υ and Ψ are right-(left-)ideals in the ring Λ, then their product $\Upsilon\Psi$ is the smallest right-(left-)ideal in Λ which contains all the products $\upsilon\psi$ for υ in Υ and ψ in Ψ. Annulets are ideals. Thus the product of two right-(left-)annulets is a well-determined right-(left-)ideal, though it need not be an annulet.

[37] For further characterizations of this class of groups, see Theorem 8.3 below and Asano (1), Satz 8, p. 243.

[38] Both finite and infinite sums.

[39] Rings with the properties (3) and (5) have been termed quasi-Frobeniusean by Nakayama (1).

THEOREM 8.1: *Suppose that G is a primary abelian operator group over the (primary) ring E, and that S and T are anti-cycles in $D(G; E)$.*
(a) *If either S is an anti-cycle of maximum index in $D(G; E)$ or T is a direct summand of G and $n(G/T) \leqq n(G/S)$, then $\mathrm{P}(S)\mathrm{P}(T) = \mathrm{P}(S + GP^{n(G/T)})$*
(b) *If both S and T are direct summands of G, but $n(G/S) < n(G/T)$, then $\mathrm{P}(S)\mathrm{P}(T)$ is not a right-annulet in $\Lambda(G; E)$.*

PROOF: If σ and τ are E-automorphisms in $\mathrm{P}(S)$ and $\mathrm{P}(T)$ respectively, then $S \leqq N(\sigma) \leqq N(\sigma\tau)$, $G^{\sigma\tau} \leqq G^{\tau}$ and G^{τ} is a cycle in $D(G; E)$ whose order is at most $n(G/T)$. Hence $n(G^{\sigma\tau}) \leqq n(G^{\tau}) \leqq n(G/T)$ so that $GP^{n(G/T)}$ is part of $N(\sigma\tau)$, since $G/N(\sigma\tau)$ is a cycle of order $n(G^{\sigma\tau})$. The product $\sigma\tau$ is therefore an element in $\mathrm{P}(S + GP^{n(G/T)})$, i.e.

$$\mathrm{P}(S)\mathrm{P}(T) \leqq \mathrm{P}(S + GP^{n(G/T)}).$$

Since S and T are anti-cycles in $D(G; E)$, there exist elements v, w such that $G = S + vE = T + wE$. Clearly $n(G/S) \leqq n(v)$, $n(G/T) \leqq n(w)$. If one of the hypotheses of (a) is satisfied by S and T, then it is possible to select the element w in such a way that $n(w) \leqq n(G/S)$, since G is the sum of T and of a cycle of order $n(G/T)$, if T is a direct summand of G and since the maximum index of the anti-cycles in $D(G; E)$ is just the maximum order of the cycles in $D(G; E)$. Consequently there exists an E-automorphism σ in $\mathrm{P}(S)$ which maps v upon w. If φ is an E-automorphism in $\mathrm{P}(S + GP^{n(G/T)})$, then $S^{\varphi} = 0$ and $n(v^{\varphi}) \leqq n(G/T)$; and we note that φ is completely determined among the automorphisms in $\mathrm{P}(S)$ by the image of v. There exists one and only one E-automorphism τ in $\mathrm{P}(T)$ which maps w upon v^{φ}; and it is readily verified that $\varphi = \sigma\tau$, proving (a).

If both S and T are direct summands of G, then we may assume that the elements v and w selected above are of order $n(G/S)$ and $n(G/T)$ respectively. Denote by ϵ the E-automorphism in $\mathrm{P}(S)$ which leaves invariant the element v. If $n(G/S) < n(G/T)$, and if σ and τ are automorphisms in $\mathrm{P}(S)$ and $\mathrm{P}(T)$ respectively, then G^{σ} is a cycle in $D(G; E)$ whose order $n(G^{\sigma})$ is at most $n(G/S) < n(G/T)$ so that $T + G^{\sigma} \leqq T + GP$. Consequently $G^{\sigma\tau} \leqq (T + GP)^{\tau} \leqq GP$ though $G^{\epsilon} \not\leqq GP$. Since $\mathrm{P}(S)$ is the smallest right-annulet containing $\mathrm{P}(S)\mathrm{P}(T)$, we have thus shown that $\mathrm{P}(S)\mathrm{P}(T)$ is not a right-annulet.

THEOREM 8.2: *Suppose that G is a primary abelian operator group over the (primary) ring E, and that S and T are cycles in $D(G; E)$.*
(a) *If either T is a cycle of maximum order in $D(G; E)$ or S is a direct summand of G and $n(S) \leqq n(T)$, then $\Lambda(S)\Lambda(T) = \Lambda(TP^{n(T)-n(S)})$.*
(b) *If both S and T are direct summands of G, but $n(T) < n(S)$, then $\Lambda(S)\Lambda(T)$ is not a left-annulet in $\Lambda(G; E)$.*

PROOF: If σ and τ are E-automorphisms in $\Lambda(S)$ and $\Lambda(T)$ respectively, then G^{σ} is a subcycle of S and $G^{\sigma\tau}$ is therefore a subcycle of T whose order does not exceed $n(S)$ so that $G^{\sigma\tau} \leqq TP^{n(T)-n(S)}$ or

$$\Lambda(S)\Lambda(T) \leqq \Lambda(TP^{n(T)-n(S)}).$$

We may assume that $S \neq 0$. There exists a greatest E-admissible subgroup

Z of G whose cross-cut with S is 0. If Z^* is an E-admissible subgroup such that $Z < Z^*$, then Z^* contains the uniquely determined subcycle $SP^{n(S)-1}$ of S; and hence it follows from **1F** that Z is an anti-cycle in $D(G; E)$. If V is an anti-cycle in $D(G; E)$ whose cross-cut with S is 0 and whose index is as small as possible, and if one of the hypotheses of (a) is satisfied by S and T, then $n(G/V) \leqq n(T)$. There exists an E-automorphism τ of G such that $V^\tau = 0$ and G^τ is the subcycle $TP^{n(T)-n(G/V)}$ of order $n(G/V)$ of T; and τ clearly belongs to $\Lambda(T)$. Suppose now that φ belongs to $\Lambda(TP^{n(T)-n(S)})$. Then G^φ is a subcycle of T whose order does not exceed $n(S)$, and $G/N(\varphi)$ is therefore a cycle of an order not exceeding $n(S)$ and is generated by some coset C. Since a coset of G/V whose order does not exceed $n(S)$ contains one and only one element in S, there exists one and only one element v in S which is mapped by τ upon C^φ. There exists furthermore an E-automorphism σ in $\Lambda(S)$ which maps $N(\varphi)$ upon 0 and C upon v; and it is readily verified that $\varphi = \sigma\tau$, proving (a).

If $n(T) < n(S)$, and if S is a direct summand of G, then it is readily verified that $\Lambda(T)$ is the smallest left-annulet which contains $\Lambda(S)\Lambda(T)$. If T is a direct summand of G, then there exists an E-automorphism ω of G which maps G upon T and which leaves invariant every element in T. This automorphism ω is in $\Lambda(T)$, but it cannot be in $\Lambda(S)\Lambda(T)$. For if σ is in $\Lambda(S)$ and τ in $\Lambda(T)$, then T^σ is a subcycle of S whose order does not exceed $n(T) < n(S)$ so that $T^\sigma \leqq SP \leqq GP$, and hence $T^{\sigma\tau} \leqq TP$, though $T^\omega = T$; and this completes the proof.

THEOREM 8.3: *If G is a primary abelian operator group over the (primary) ring E, then each of the following properties implies the others*:

(a) *G is the direct sum of a finite number of cycles of equal order in $D(G; E)$.*
(b) *The product of any two right-annulets in $\Lambda(G; E)$ is a right-annulet.*
(c) *The product of any two left-annulets in $\Lambda(G; E)$ is a left-annulet.*

PROOF: That (a) implies both (b) and (c), is an immediate consequence of Theorem 7.5. If (b) or (c) is satisfied by G, then we deduce from Theorems 8.1, (b) and 8.2, (b) that G is the direct sum of cycles of equal order in $D(G; E)$, since G is, by **1E**, a direct sum of cycles in $D(G; E)$. If Z is any cycle of maximum-order in $D(G; E)$, then Z is a direct summand of G and there exists to every element g in G an E-automorphism of G which maps Z upon the cycle gE. Hence it follows from Lemma 6.2 that $\Lambda(Z)\Lambda(G) = \Lambda(Z)\mathrm{A} = \Phi$, the ideal of all the finite automorphisms. Thus (c) implies that **1** is a finite E-automorphism, showing that (a) is a consequence of (c). If V is an anti-cycle of maximum-index in $D(G; E)$, then V is a direct summand of G and every anti-cycle in $D(G; E)$ may be mapped upon V by an E-automorphism of G. Hence we infer from Lemma 6.2 that $\mathrm{P}(0)\mathrm{P}(V) = \mathrm{A}\mathrm{P}(V) = \Phi$. Consequently **1** is a finite E-automorphism, if (b) is satisfied by G; and (a) is therefore an implication of (b).

9. Cross-cuts of Right- and Left-Annulets

Right- and left-annulets in A are ideals in A; and they are therefore subrings of A. The cross-cut[40] of a right-annulet and a left-annulet in A is an ideal in

[40] The cross-cut of the sets S' and S'' shall always be denoted by $S' \cap S''$.

both these subrings of A. It is the object of this section to investigate the ideal theory obtained this way.

THEOREM 9.1: *Suppose that G is a primary abelian operator group over the (primary) ring E, and that H is an anti-cycle of index $n \neq 0$ in $D(G; E)$.*

(a) *If H is a direct summand of G, then every left-ideal I in the ring $\mathrm{P}(H)$ is the cross-cut of $\mathrm{P}(H)$ and of a left-ideal in $\mathrm{A}(G; E)$.*

(b) *If I is the cross-cut of $\mathrm{P}(H)$ and of some left-ideal in $\mathrm{A}(G; E)$, then there exists one and only one E-admissible subgroup S of G such that the orders of the elements in S do not exceed n and such that $\mathrm{I} = \mathrm{P}(H) \cap \mathrm{A}(S)$.*

(c) *If the orders of the elements in the E-admissible subgroups S_v of G do not exceed n, then $\sum_v (\mathrm{P}(H) \cap \mathrm{A}(S_v)) = \mathrm{P}(H) \cap \mathrm{A}(\sum_v S_v)$.*

PROOF: If H is a direct summand of G, then there exists an (idempotent) E-automorphism η of G which maps H upon 0 and every coset C of G/H upon an element in C. If I* is the smallest left-ideal in A containing I, then every element in I* is the sum of elements $\varphi\chi$ for φ in A and χ in I. If ρ is an element in $\mathrm{P}(H)$, then $\rho = \eta\rho$; and thus it follows that the sum of elements $\varphi\chi$, if in $\mathrm{P}(H)$, is equal to the sum of elements $(\eta\varphi)\chi$. Since $\eta\varphi$ is in $\mathrm{P}(H)$, and since I is a left-ideal in the ring $\mathrm{P}(H)$, it follows that every $(\eta\varphi)\chi$ belongs to I and that therefore[41] $\mathrm{I} = \mathrm{I}^* \cap \mathrm{P}(H)$.

If I is the cross-cut of $\mathrm{P}(H)$ and of some left-ideal in A, then I is in particular the cross-cut of $\mathrm{P}(H)$ and I*, the smallest left-ideal in A containing I. Since every element in I induces a linear transformation of the cycle G/H of order n upon some subcycle of G, it follows that the orders of the elements in $S = G^{\mathrm{I}}$ do not exceed n. If C is some coset which generates the cycle G/H, then S consists of all the elements $C^{\chi}e$ for χ in I and e in E. It is a consequence of Lemma 5.1 that there exists an E-automorphism φ of G which maps H into part of H and the coset C into part of Ce. Since $C^{\chi\varphi} = C^{\chi}e$, and since $\varphi\chi$ is in I* and maps H upon 0, it follows that $\varphi\chi$ is in I and that S consists exactly of all the elements C^{ι} for ι in I. Thus S is an E-admissible subgroup of G the orders of whose elements do not exceed n. If κ is an element in $\mathrm{P}(H) \cap \mathrm{A}(S)$, then there exists an element ζ in I such that $C^{\zeta} = C^{\kappa}$, since S consists of all the C^{ι} for ι in I. But this implies clearly $\zeta = \kappa$ so that $\mathrm{I} = \mathrm{P}(H) \cap \mathrm{A}(S)$. Suppose now that T is an E-admissible subgroup of G the orders of whose elements do not exceed n and which satisfies: $\mathrm{I} = \mathrm{P}(H) \cap \mathrm{A}(T)$. If t is an element in T, then $n(tE) = n(t) \leqq n = n(G/H)$; and it is a consequence of **1E** that there exists an E-automorphism γ of G which maps H upon 0 and C upon t (γ clearly belongs to I). Hence $T = G^{\mathrm{I}} = S$.

If the orders of the elements in the E-admissible subgroups S_v of G do not exceed n, then the orders of the elements in the E-admissible subgroup $S = \sum_v S_v$ do not exceed n (by **1A**). Clearly $\sum_v (\mathrm{P}(H) \cap \mathrm{A}(S_v)) \leqq \mathrm{P}(H) \cap \mathrm{A}(S)$. If φ is an element in $\mathrm{P}(H) \cap \mathrm{A}(S)$, then C^{φ} for C a generating coset of G/H is an element in S so that $C^{\varphi} = s_1 + \cdots + s_k$ where s_i is an ele-

---

[41] The hypothesis that H is an anti-cycle in $D(G;E)$ has not been used in this part of the proof.

ment in S_{v_i}. There exists an E-automorphism φ_i in $\mathrm{P}(H) \cap \Lambda(S_{v_i})$ which maps C upon s_i. Clearly $\varphi = \varphi_1 + \cdots + \varphi_k$, proving the desired equality (c).

THEOREM 9.2: *Suppose that G is a primary abelian operator group over the (primary) ring E, that Z is a cycle of order $n \neq 0$ in $D(G; E)$, and that* I *is a right-ideal in the subring $\Lambda(Z)$ of $\Lambda(G; E)$.*

(a) *If Z is a direct summand of G, then* I *is the cross-cut of $\Lambda(Z)$ and of a right-ideal in Λ.*[42]

(b) *There exists at most one E-admissible subgroup S of G such that the orders of the elements in G/S do not exceed n and such that $\mathrm{I} = \Lambda(Z) \cap \mathrm{P}(S)$.*

(c) *If* I *is the cross-cut of $\Lambda(Z)$ and of a right-ideal in Λ, and if $G/N(\mathrm{I})$ is a sum of a finite number of cycles, then the orders of the elements in $G/N(\mathrm{I})$ do not exceed n and $\mathrm{I} = \Lambda(Z) \cap \mathrm{P}(N(\mathrm{I}))$.*

PROOF: If Z is a direct summand of G, then there exists an (idempotent) E-automorphism ζ of G which maps G upon Z and which leaves invariant all the elements in Z. The smallest right-ideal in Λ which contains I is $\mathrm{I}^* = \mathrm{I}\Lambda$, and every element in I^* is the sum of products $\chi\varphi$ for χ in I and φ in Λ. If ω is an element in $\Lambda(Z)$, then $\omega = \omega\zeta$; and every element in $\mathrm{I}^* \cap \Lambda(Z)$ is therefore the sum of products $\chi(\varphi\zeta)$ for χ in I and φ in Λ. Since every $\varphi\zeta$ is in $\Lambda(Z)$, and since I is a right-ideal in $\Lambda(Z)$, it follows that every $\chi\varphi\zeta$ is in I, i.e. $\mathrm{I} = \Lambda(Z) \cap \mathrm{I}^*$.

The statement (b) is an immediate consequence of the following lemma.

(b*) If S is an E-admissible subgroup of G, if the orders of the elements in G/S do not exceed n, then $S = N(\Lambda(Z) \cap \mathrm{P}(S))$.

It is clear that $S \leqq N(\Lambda(Z) \cap \mathrm{P}(S))$. To prove the opposite inequality consider a basis B of G/S. Since the orders of the elements in B do not exceed n, there exists to every element b in B an E-automorphism $\varphi(b)$ of G in $\mathrm{P}(S)$ which maps the coset b of G/S upon an element of order $n(b)$ in Z and which maps all the other elements in B upon 0. If x is an element not 0 in G/S, then $x = \sum_{b \text{ in } B} be(b)$ where the $e(b)$ are elements in E almost all of which are 0. There exists at least one element in B, say v, such that $ve(v) \neq 0$; and it is clear that $x^{\varphi(v)}$ is different from 0. This completes the proof of (b^*) (and (b)).

If γ is any automorphism in $\Lambda(Z)$, then $G/N(\gamma)$ is a cycle of an order not exceeding n and consequently $GP^n \leqq N(\gamma)$. If I is any subset of $\Lambda(Z)$, then $N(\mathrm{I})$ is the cross-cut of all the $N(\gamma)$ for γ in I so that $GP^n \leqq N(\mathrm{I})$; and the orders the of elements in $G/N(\mathrm{I})$ do not exceed n.

If I is the cross-cut of $\Lambda(Z)$ and of some right-ideal in Λ, then $\mathrm{I} = \Lambda(Z) \cap \mathrm{I}^*$ where $\mathrm{I}^* = \mathrm{I}\Lambda$ is the smallest right-ideal in Λ which contains I. It is clear that $N(\mathrm{I}) = N(\mathrm{I}^*)$ and that $\mathrm{I} \leqq \Lambda(Z) \cap \mathrm{P}(N(\mathrm{I}))$. We consider the set Q of all the pairs (U, V) with the following properties:

(i) U and V are E-admissible subgroups of $G/N(\mathrm{I})$ and $G/N(\mathrm{I})$ is the direct sum of U and V.

(ii) To every automorphism $\mathfrak{s}$ in $\Lambda(Z) \cap \mathrm{P}(N(\mathrm{I}))$ there exists an automorphism $\mathfrak{s}^*$ in I which maps V upon 0 and which coincides with $\mathfrak{s}$ on U.

---

[42] For the proof of (a) we need not assume that Z is a cycle in $D(G;E)$.

This set Q is not vacuous, since the pair $(0, G/N(\mathrm{I}))$ meets the requirements (i), (ii). To prove (c) we assume that $G/N(\mathrm{I})$ is the sum of a finite number of cycles. Thus the maximum condition is satisfied by the E-admissible subgroups of $G/N(\mathrm{I})$ and hence there exists a pair (U^*, V^*) in Q whose first member U^* is as big as possible To prove (c) we need only show that $V^* = 0$. If V^* were not 0, then V^* would contain a subcycle W of maximum order in V^*. To every element, not 0, in W there exists an automorphism in I which maps this element upon an element, not 0, in Z (as follows from the definition of $N(\mathrm{I})$ and from the fact that $\mathrm{I} \leqq \Lambda(Z)$). Thus there exists an automorphism δ in I which maps the cycle $WP^{n(W)-1}$ of order 1 upon the cycle ZP^{n-1} of order 1; and it is readily seen that $W^\delta = ZP^{n-n(W)}$ is a cycle of order $n(W)$. Since (U^*, V^*) belongs to Q, there exists an automorphism δ^* in I which coincides with δ in U^* and which maps V^* upon 0. The automorphism $\epsilon = \delta - \delta^*$ belongs to I (since I is an ideal), maps U^* upon 0 and W upon $ZP^{n-n(W)}$. Denote by V^{**} the set of elements in V^* which are mapped upon 0 by ϵ, and put $U^{**} = U^* + W$. Since ϵ maps V^* upon a cycle of order $n(W)$, and since W is a subcycle of maximum order of V^*, it follows that $G/N(\mathrm{I})$ is the direct sum of the E-admissible subgroups U^{**} and V^{**}. If φ is any automorphism in $\Lambda(Z) \cap \mathrm{P}(N(\mathrm{I}))$, then there exists—by (ii)—an automorphism φ^* in I which coincides with φ on U^* and which maps V^* upon 0. Since φ maps W upon part of $ZP^{n-n(W)} = W^\epsilon$, there exists an E-automorphism ρ of $ZP^{n-n(W)}$ such that φ and $\epsilon\rho$ coincide on W. There exists by Lemma 5.1 an E-automorphism σ of G which induces ρ in $ZP^{n-n(W)}$; and $\epsilon\sigma$ is clearly an automorphism in $\mathrm{I} = \mathrm{I}^* \cap \Lambda(Z)$. The automorphism $\varphi^{**} = \varphi^* + \epsilon\sigma$ belongs to I, maps V^{**} upon 0 and coincides with φ on U^{**}. Thus we have shown that the pair (U^{**}, V^{**}) belongs to the set Q in spite of the fact that $U^* < U^{**}$, a contradiction which completes the proof.

COROLLARY 9.3: *Suppose that G is a primary abelian operator group over the (primary) ring E, and that Z is a cycle of order $n(\neq 0)$ in $D(G; E)$.*
(a) *G/GP^n is the sum of a finite number of cycles if, and only if, $\Lambda(Z) \cap \mathrm{I} = \Lambda(Z) \cap \mathrm{P}(N(\Lambda(Z) \cap \mathrm{I}))$ for every right-ideal I in $\Lambda(G; E)$.*
(b) *If S and T are E-admissible subgroups of G such that the orders of the elements in G/S and in G/T do not exceed n, then $(\Lambda(Z) \cap \mathrm{P}(S)) + (\Lambda(Z) \cap \mathrm{P}(T)) = \Lambda(Z) \cap \mathrm{P}(S \cap T)$.*

PROOF: $GP^n \leqq N(\Lambda(Z) \cap \mathrm{I})$, since Z is a cycle of order n. If G/GP^n is the sum of a finite number of cycles, then $G/N(\Lambda(Z) \cap \mathrm{I})$ is the sum of a finite number of cycles; and $\Lambda(Z) \cap \mathrm{I} = \Lambda(Z) \cap \mathrm{P}(N(\Lambda(Z) \cap \mathrm{I}))$ for right-ideals I in Λ is a consequence of Theorem 9.2, (c). Suppose now that this identity holds for every right-ideal I in Λ. It is a consequence of **1E** that G/GP^n is the direct sum of cycles Z_v of positive order not exceeding n. The set I of all the E-automorphisms which map GP^n and almost all the Z_v upon 0 is readily seen to be a right-ideal in Λ; and the cross-cut $\mathrm{I} \cap \Lambda(Z)$ consists exactly of those E-automorphisms of G which map G into part of Z and which map GP^n and almost all the Z_v upon 0. Since $n(Z_v) \leqq n = n(Z)$, it follows that $N(\mathrm{I}) = GP^n =$

$N(\Lambda(Z) \cap \mathrm{I})$. Hence there exists an automorphism ω in $\mathrm{P}(N(\Lambda(Z) \cap \mathrm{I}))$ which maps GP^n upon 0 and [every] Z_v upon $ZP^{n-n(Z_v)}$. Since ω belongs to $\Lambda(Z)$, it follows from the validity of the identity in (a), that ω belongs to I so that the automorphism ω which maps none of the Z_v upon 0 maps almost all the Z_v upon 0. Hence there exists only a finite number of Z_v ; and this completes the proof of (a).

To prove (b) we need only show that every automorphism φ in $\Lambda(Z) \cap \mathrm{P}(S \cap T)$ belongs to $(\Lambda(Z) \cap \mathrm{P}(S)) + (\Lambda(Z) \cap \mathrm{P}(T))$. Since the orders of the elements in G/S and G/T do not exceed n, it follows that GP^n is part of S and of T, and that therefore the orders of the elements in $G/(S \cap T)$ do not exceed n. Hence an E-automorphism γ of G may be constructed which maps S upon 0 and coincides with φ on T and which maps G into part of Z. Then τ is in $\Lambda(Z) \cap \mathrm{P}(S)$ and $\varphi - \tau$ is in $\Lambda(Z) \cap \mathrm{P}(T)$; and this completes the proof of (b).

REMARK: In sections **5** to **8** we obtained fairly similar results for right- and for left-annulets, though we had to use different methods in proving these facts. In this section **9** even the results turned out to be different (cf. in particular Corollary 9.3, (a) and Theorem 9.1, (c)). The following argument will give some explanation of this difference in behaviour between right- and left-annulets.

1. If Z is a cycle of order $n \neq 0$ in $D(G; E)$, E a primary ring, then GP^n is mapped upon 0 by all the automorphisms in $\Lambda(Z)$; and the automorphisms in $\Lambda(Z)$ map G/GP^n into the cycle Z. Thus the automorphisms in $\Lambda(Z)$ behave just like the characters of G/GP^n with values in Z; and $\Lambda(Z)$ is essentially the character group of G/GP^n (in Z).[43] If S is an E-admissible subgroup of G which contains GP^n, then $\Lambda(Z) \cap \mathrm{P}(S)$ consists just of those characters of G/GP^n which map S upon 0. This correspondence between the subgroups in $D(G/GP^n; E)$ and right-ideals in the ring $\Lambda(Z)$ is a duality of $D(G; E)$ upon a certain system of right-ideals in the ring $\Lambda(Z)$.

2. If H is an anti-cycle of index $n \neq 0$ in $D(G; E)$, C a generating coset of the cycle G/H, and G_n the subgroup of all the elements in G whose order does not exceed n, then there exists to every element x in G_n one and only one E-automorphism $\varphi = \varphi(x)$ in $\mathrm{P}(H)$ such that $x = C^\varphi$, and hence $C^{\mathrm{P}(H)} = G_n$. Mapping x upon $\varphi(x)$ constitutes an isomorphism of G_n upon $\mathrm{P}(H)$; and $\mathrm{P}(H) \cap \Lambda(S)$ is just the image of the E-admissible subgroup S of G under this isomorphism. Thus G_n and $\mathrm{P}(H)$ are essentially the same.

3. If the group G is the sum of a finite number of cycles, then it is well known[44] that G and its character group are very similar and their respective systems of subgroups are duals. Without this hypothesis this duality breaks down[45] and can only be reconstructed by resort to topological means.[46] Since these do not

[43] Cp. Lewis (1).

[44] Cp. e.g. Baer (4), II, 4 and Lewis (1).

[45] It has been shown that this finiteness-hypothesis is (under certain circumstances) actually necessary for the existence of a dual; cp. Baer (3).

[46] Pontrjagin (1), Ch. V.

enter into our considerations, G and its character group will in general be rather dissimilar; and thus right- and left-annulets will behave differently.

Chapter IV: Characterization of Automorphism Rings

In Chapter III we have derived a number of properties which are satified by the annulets in the automorphism rings of primary abelian operator groups. It is the object of this present Chapter IV to prove that the automorphism rings of primary abelian operator groups are the only rings with these properties.

10. Idempotents

We consider a ring K which shall always be assumed to contain an identity element 1. If S is any subset of K, then we define as the right (resp. left)-annulet $R(S)$ (resp. $L(S)$) determined by S the set of all the elements x in K which satisfy: $sx = 0$ (resp. $xs = 0$) for every s in S. The right (resp. left)-annulets are right (resp. left)-ideals; cross-cuts of right (resp. left)-annulets are right (resp. left)-annulets; $S \leqq T$ implies $R(T) \leqq R(S)$ (resp. $L(T) \leqq L(S)$); and $R(L(R(S))) = R(S)$ (resp. $L(R(L(S))) = L(S)$).

If U and V are subsets of K, then their sum $U + V$ consists of all the sums $u + v$ for u in U and v in V. If U and V are ideals whose cross-cut is 0, then $U + V$ is the direct sum of U and V, and U is a direct summand of $U + V$. An element g in K is an idempotent if $g = g^2$. In the following we collect certain well known[47] connections between the concepts: annulet, direct summand of the ring K and idempotent, which will prove useful in the future; but it should be kept in mind that these remarks refer only to an important subclass of the set of annulets.

Lemma 10.1: *The ring K is the direct sum of the right (resp. left)-ideals U and V if, and only if, there exist idempotents u and v such that $1 = u + v$, $uv = vu = 0$ and $U = uK$, $V = vK$ (resp. $U = Ku$, $V = Kv$).*

Lemma 10.2: *If the idempotents u and v in the ring K satisfy: $1 = u + v$ and $uv = vu = 0$, then $R(u) = R(Ku) = vK$, $R(v) = R(Kv) = uK$ (resp. $L(u) = L(uK) = Kv$, $L(v) = L(vK) = Ku$).*

Lemma 10.3: *The right (resp. left)-ideal U is a direct summand of K if, and only if, there exists an idempotent u such that $U = uK$ (resp. $U = Ku$).*

Lemma 10.4: *If u is an idempotent and S a subset of the ring K which contains uS and/or Su, then $uS = S \cap uK$ and/or $Su = S \cap Ku$.*

We note in particular the following important consequence of Lemmas 10.1 and 10.2: If K is the direct sum of the right (resp. left)-ideals U and V, then U and V are right (resp. left)-annulets whose cross-cut is 0; and K is the direct sum of $L(U)$ and $L(V)$ (resp. $R(U)$ and $R(V)$) so that $L(U) \cap L(V) = 0$ (resp. $R(U) \cap R(V) = 0$).

11. Groups Contained in the Ring K

If J is a right-ideal in the ring K, then J is a subring of K and in particular J

[47] Cp. e.g. von Neumann (1), Ch. II, p. 7/8.

is an additive abelian group. Every cross-cut of J and of a left-ideal in K is a subgroup of this abelian group J and *the set of all these subgroups* shall be denoted by $D(J)$.

If x is any element in K satisfying $xJ \leqq J$, then $xS \leqq S$ for every subgroup S of the group J which is contained in $D(J)$; the elment x induces in J an endomorphism (cf. section 4) which may be termed *$D(J)$-admissible*, since it maps every subgroup in $D(J)$ into itself. *The system of all the endomorphisms of J which are obtained by left-multiplication by elements in K* shall be denoted by $E(J)$. The elements x in K which satisfy: $xJ \leqq J$ form a subring $L^*(J)$ of K which clearly contains $L(J)$; and one verifies readily that mapping the element x in $L^*(J)$ upon the endomorphism in $E(J)$ which x induces in J constitutes an anti-isomorphism of $L^*(J)/L(J)$ upon $E(J)$.

If x is any element in K, then $Jx \leqq J$, since J is a right-ideal in K and the element x induces therefore in this fashion an endomorphism in the group J. If e is any endomorphism in $E(J)$, then e maps every element y in J upon an element ye in J; and there exists by the definition of $E(J)$ an element f in $L^*(J)$ such that $fy = ye$ for every y in J. If x is any element in K, y an element in J, e an element in $E(J)$ and f an element in $L^*(J)$ inducing e in J, then $(ye)x = (fy)x = f(yx) = (yx)e$ so that *x induces in J an $E(J)$-automorphism* which we denote by y^x. *The system of these $E(J)$-automorphisms of J* shall be denoted by $A(J)$; and it is clearly isomorphic to $K/R(J)$ (this is not important, since in all our applications $R(J)$ will be 0).

LEMMA 11.1: *If the right-ideal J in K is generated by some idempotent j, i.e. if $J = jK$, then every endomorphism in $E(J)$ is induced by one and only one element in the ring jKj (with identity element j) and in this fashion an anti-isomorphism of the ring jKj upon the ring $E(J)$ is defined; and $D(J)$ is exactly the system $D(J; E(J))$ of all the $E(J)$-admissible subgroups of J as well as the system of all the left-ideals in the ring J.*

PROOF: An element y belongs to the right-ideal $J = jK$, for j an idempotent, if, and only if, $y = jy$. If e is any element in $E(J)$, then there exists an element f in $L^*(J)$ such that $ye = fy$ for every y in J. Since y and ye are elements in J, we have $y = jy$ and $ye = j(ye)$; and thus it follows that $ye = j(ye) = j(fy) = (jf)y = (jf)(jy) = (jfj)y$. It is obvious that $jKj \leqq L^*(J)$ and that jfj induces the null-element of $E(J)$ if, and only if, $jfj = 0$.

The cross-cut of J and of a left-ideal in K is a left-ideal in the ring J. From what we have shown just now it follows that every left-ideal in the ring J is an $E(J)$-admissible subgroup of the abelian group J. If finally S is an $E(J)$-admissible subgroup of the abelian group J, then $S = jS$, since $S \leqq jK = J$, and $S = SE(J)$. Consequently we may deduce from Lemma 10.4 that the cross-cut of the left-ideal KS and of J is exactly $KS \cap J = jKS = jKjS = SE(J) = S$; and this completes the proof of our lemma.

12. The Postulates

In this section we are going to enumerate a number of properties of rings; we shall show that the automorphism rings of primary abelian operator groups

meet these requirements and we shall discuss a few of the interrelations between these postulates. In the next section we shall prove that the automorphism rings of primary abelian operator groups are the only rings satisfying these conditions.

If K is a ring (containing an identity element 1), then we shall denote by $RA(K)$ (resp. $LA(K)$) *the partially ordered set of all the right (resp. left)-annulets in* K. Both $RA(K)$ and $LA(K)$ contain the (set-theoretical) cross-cuts of all their subsets; and thus they contain with any subset a smallest annulet containing all the annulets in the given set: the *join* of the annulets in the set. We note that these joins of annulets need not be their ideal-theoretical sums. The mappings: H upon $L(H)$ and J upon $R(J)$ constitute reciprocal dualities of $RA(K)$ upon $LA(K)$ and of $LA(K)$ upon $RA(K)$.

I. *If* $H_i = R(J_i)$ *and* $J_i = L(H_i)$, *and if* $H_1 \cap H_2 = 0 = J_1 \cap J_2$, *then* $H_1 + H_2 = K = J_1 + J_2$.

It is a consequence of Theorem 7.1, (c) that this postulate I is satisfied by the automorphism rings Λ of primary abelian operator groups.

II. *If H is a cycle of order $n \neq 0$ in $RA(K)$, then*

(a) *$D(H)$ is a ring*[48] *of subgroups of the additive abelian group H; and*

(b) *to every S in $D(H)$ there exists one and only one S^* in $LA(K)$ such that $S = H \cap S^*$ and such that the orders of the subcycles of S^* in $LA(K)$ do not exceed n.*

It is a consequence of Theorem 9.1, (c), (b) that this postulate II is satisfied by the automorphism rings of primary abelian operator groups.

III. *If H is a cycle of maximum order m in $RA(K)$, if J is a cycle of order $n(\leqq m)$ in $LA(K)$, if K is the direct sum of J and some left-annulet in K, and if V is a right-ideal in K such that $V \cap J \leqq H$, then there exists a right-annulet W in K such that $W \cap J = V \cap J$.*

That III is satisfied by the automorphism rings $\Lambda(G; E)$ of primary abelian operator groups G over (primary) rings E, may be seen as follows: It is a consequence of the Fundamental Theorem of section **2** that H is a cycle of maximum order m in $RA(\Lambda)$ if, and only if, $\mathrm{H} = \mathrm{P}(S)$ for S an anti-cycle of maximum index m in $D(G; E)$; and that I is a cyclic direct summand of Λ in $LA(\Lambda)$ if, and only if, $\mathrm{I} = \Lambda(T)$ for T a cyclic direct summand of G in $D(G; E)$. If Ψ is a right-ideal in Λ such that $\Psi \cap \mathrm{I} \leqq \mathrm{H}$, then $N(\mathrm{H}) \leqq N(\Psi \cap \mathrm{I})$. Since $S = N(\mathrm{P}(S)) = N(\mathrm{H})$, it follows that $N(\Psi \cap \mathrm{I})$ is an anti-cycle in $D(G; E)$; and that therefore by Theorem 9.2, (c) we have $\Psi \cap \mathrm{I} = \Lambda(T) \cap \mathrm{P}(N(\Psi \cap \mathrm{I})) = \mathrm{P}(N(\Psi \cap \mathrm{I})) \cap \mathrm{I}$, as was to be shown.

IV, i. *There exists an integer m such that both $RA(K)$ and $LA(K)$ contain at least i independent cycles of order m and such that neither contains cycles of an order exceeding m.*

This condition IV, i shall only be used for $i = 1, 2, 3$. We note that condition IV, 1 states only the following fact: The orders of the cycles in $RA(K)$ and in $LA(K)$ are bounded and the maximum order of the cycles in $RA(K)$ is the same as the maximum order of the cycles in $LA(K)$. It is a consequence of **1E** and

[48] In the sense of **1B**.

of the Fundamental Theorem of section **2** that postulate IV, 1 is satisfied by the automorphism rings of primary abelian operator groups. It should be noticed that III would be vacuous, if IV, 1 were not true.

THEOREM 12.1: (1) *If* II, (b) *and* IV, 1 *are satisfied by the ring* K, *and if* G *is a cycle of maximum order* m *in* $RA(K)$, *then a projection of* $LA(K)$ *upon* $D(G)$ *is effected by mapping the left-annulet* J *upon* $J \cap G$.
(2) *If* II *and* IV, 1 *are satisfied by the ring* K, *then* $RA(K)$ *and* $LA(K)$ *are complete modular lattices.*

PROOF: Every element in $D(G)$ has the form (by definition of $D(G)$): $G \cap V$ for V a left-ideal in K. Since m is the maximum order of the cycles in $LA(K)$, there exists to every left-ideal V in K one and only one left-annulet V^* such that $G \cap V = G \cap V^*$; and this proves (1). (2) is an immediate consequence of (1), since rings of subgroups of abelian groups are complete modular lattices, and since $RA(K)$ and $LA(K)$ are duals of each other

V. *The following conditions are satisfied by the partially ordered set* $LA(K)$.
(a) *The modular (Dedekind's) law.*
(b) *If* S *is a not vacuous set of cycles in* $LA(K)$, *if* S *contains every cycle in* $LA(K)$ *which is part of the join of a finite number of cycles in* S, *then there exists one and only one left-annulet* L_S *in* K *such that the cycles in* S *and only these are part of* L_S.
(c) *If* U *and* V *are left-annulets in* K, *if* $U \leqq V$, *and if there exist at most two different left-annulets between* U *and* V *which are modulo*[49] U *cycles of order* 1, *then* V *is a cycle modulo* U.

It is a consequence of **1E**, **1F** and the Fundamental Theorem of section **2** that this postulate is satisfied by the automorphism rings of primary abelian operator groups.

It has already been shown (Theorem 12.1, (2)) that V, (a) may be deduced from II and IV, 1. It is easily deduced from V, (b) that *every left-annulet in* K *is the join of cycles*[50] *in* $LA(K)$; and it follows from Theorem 12.1, (1) that V, (*b*) is a consequence of II, IV, 1 and this last condition.

The part of IV, 1 which states that the maximum order of the cycles is the same in $RA(K)$ and in $LA(K)$ may be derived from V, since $LA(K)$ and $RA(K)$ are duals of each other, and since it may be shown[51] that K is in $RA(K)$ and in $LA(K)$ the direct sum of a cycle of maximum order and of an anti-cycle of maximum index.

THEOREM 12.2: *If* I, II, (*b*), IV, 1 *and* V *are satisfied by the ring* K, *and if* G *is a cycle of maximum order* m *in* $RA(K)$, *then* $G = jK$ *for some idempotent* j *in* K, *and* $D(G)$ *is a ring of subgroups of* G.

PROOF: It is known[51] that there exists because of V an element H in $RA(K)$ satisfying: K is the join and 0 the cross-cut of G and H. This implies that 0 is the cross-cut of $L(G)$ and $L(H)$; and hence it follows from I that K is the

---

[49] U is modulo V a cycle of order n, if V is part of U, and if U is a cycle of order n in the partially ordered set of the elements between U and V.

[50] Though it need not be the join of a finite number of cycles.

[51] Baer (4), Theorem I.3.6 and I.5.1.

direct sum of G and H. Now we deduce from Lemma 10.1 that $G = jK$ for j an idempotent in K; and we deduce from Lemma 11.1 that $D(G)$ is just the ring of all the $E(G)$-admissible subgroups of the abelian group G.

VI. *If X and Y are cycles in $RA(K)$, and if X is a cycle of maximum order in $RA(K)$, then*[52] $XY = X^{(n(X)-n(Y))}$.

It is a consequence of Theorem 8.1, (a) that VI is satisfied by the automorphism rings of primary abelian operator groups.

The last postulate which we require will be vacuous, if K is the join of a finite number of cycles in $RA(K)$. To enunciate postulate VII two concepts are needed: We denote by $W = W(K)$ *the class of all the right-annulets in K which are cross-cuts of a finite number of anti-cycles in $RA(K)$*; and we term *W-function* any single-valued function $f(J)$ of the right-annulets J in the class W with the following properties:

$$f(J) \text{ is a coset of } K/J; \quad \text{and } J \leqq H \text{ implies } f(J) \leqq f(H).$$

We note that 0 is in the class W if, and only if, 0 is the cross-cut of a finite number of anti-cycles in $RA(K)$; and this is equivalent to saying that K is the join of a finite number of cycles in $RA(K)$.

VII. *To every W-function $f(J)$ there exists one and only one element f in K such that $f(J) = J + f$ for every J in W.*

It is a consequence of Theorem 6.3 that the postulate VII is satisfied by the automorphism rings of primary abelian operator groups.

It may be mentioned that the uniqueness of the element f in K occurring in VII may be derived from the other postulates.

13. Completeness of the System of Postulates

In this section we are going to prove the following *Existence Theorem*:

The postulates I to III, IV, 2, and V to VII are satisfied by the ring K containing an identity element 1 if, and only if, there exists a primary abelian operator group G over a (primary) ring E such that $D(G;E)$ contains at least two independent cycles of maximum order and such that K is isomorphic to the ring $A(G;E)$ of all the E-automorphisms of G.

It has been shown in section **12** that the postulates I to III, IV, 1, V to VII are satisfied by the automorphism rings of primary abelian operator groups; and that IV, 2 is satisfied whenever $D(G;E)$ contains at least two independent cycles of maximum order, is an immediate consequence of the Fundamental Theorem of section **2**.

If IV, 1 is satisfied by the ring K, then there exists a cycle G of maximum order m in the set $RA(K)$ of all the right-annulets in K. We are going to prove that

G is a primary abelian operator group over the (primary) ring $E(G)$, that $D(G;E(G))$ contains at least two independent cycles of the maximum order m and

[52] For these notations cp. **1A** and section **8**.

that K is (essentially) the ring of all the $E(G)$-automorphisms of the abelian group G, provided I, II, b, III, IV, 2, V to VII are satisfied by the ring K.

The proof of this theorem will be effected in several steps; and we shall make a note at each of these lemmas which of the hypotheses we actually applied in its proof.

13.1. *[$G = jK$ for some idempotent j in K;] $D(G)$ is the ring of all the left-ideals in the ring G and $D(G) = D(G;E(G))$; every endomorphism in $E(G)$ is induced by left-multiplication of G by one and only one element in the ring jKj; the rings jKj and $G/(L(G) \cap G)$ are isomorphic and the rings $E(G)$ and jKj are anti-isomorphic.* (I; II, b; IV, 1; V.)

This statement is an immediate consequence of Theorem 12.2 and of Lemma 11.1.

G is a cycle of order m in $RA(K)$ and hence it contains one and only one subcycle $G^{(i)}$ of order $m - i$ in $RA(K)$ for $0 \leqq i \leqq m$.

13.2. *$G^{(i)}$ is a two-sided ideal in the subring G of K* (VI).

Proof: $G^{(i)}$ is a right-annulet and therefore a right-ideal in K. Since G is a cycle of maximum order m in $RA(K)$, it follows from VI that $GG^{(i)} \doteq G^{(i)}$ so that $G^{(i)}$ is a left-ideal in the ring G.

13.3. *$G^{(m-i)}$ is the sum of all the cycles of an order not exceeding i in $D(G)$* (I; II, b; IV, 1; V; VI).

Proof: Since $G^{(m-i)}$ is a cycle of order i in $RA(K)$, and since $G^{(m-i)}$ belongs to $D(G^{(m-i)})$, there exists one and only one left-annulet L in K which contains $G^{(m-i)}$ and which contains only cycles in $LA(K)$ of an order not exceeding i. If C is any cycle in $LA(K)$ whose order does not exceed i, then the orders of the subcycles of the join of L and C in $LA(K)$ do not exceed i (by **1A**). Hence C is part of L, since $G^{(m-i)}$ is part of the join of L and C. Clearly L is the smallest left-annulet containing $G^{(m-i)}$; and it is a consequence of V, b that L is the join of all the cycles in $LA(K)$ whose orders do not exceed i.

It is a consequence of 13.2 and 13.1 that $G^{(m-i)}$ belongs to $D(G)$. Hence there exists by II, b one and only one left-annulet L^* such that $G^{(m-i)} = G \cap L^*$. Since L is the smallest left-annulet containing $G^{(m-i)}$, this implies $L = L^*$. It is a consequence of II, b that a projectivity of $LA(K)$ upon $D(G)$ is obtained by mapping every left-annulet in K upon its cross-cut with G. Hence we may deduce from 13.1 that $G^{(m-i)} = G \cap L$ is the sum of all the cycles in $D(G)$ of an order not exceeding i.

13.4. *$E(G)$ is a cycle of order m in the partially ordered set of all the right-ideals[53] in the ring $E(G)$* (I; II, b; IV, 1; V).

Proof: It is a consequence of 13.1 that the rings $G/(L(G) \cap G)$ and $E(G)$ are anti-isomorphic; and hence it suffices to prove that G is a cycle of order m in the partially ordered set of left-ideals in G which contain $L(G) \cap G$. If J is such a left-ideal in G, then it follows from 13.1 that J belongs to $D(G)$; and thus we may deduce from II, c that there exists one and only one left-annulet L

[53] This implies that every right-ideal in $E(G)$ is a two-sided ideal.

satisfying $J = G \cap L$. It is a consequence of II, b and of $L(G) \cap G \leqq L \cap G$, that $L(G) \leqq L$; and this implies $R(L) \leqq R(L(G)) = G$, since G is a right-annulet. But the $G^{(i)}$ for $0 \leqq i \leqq m$ are the only right-annulets contained in G so that $R(L) = G^{(i)}$, $L = L(G^{(i)})$ and $J = L(G^{(i)}) \cap G$ for some i. G is therefore a cycle of an order not exceeding m in the partially ordered set of left-ideals in G which contain $L(G) \cap G$. The order is exactly m, since $G^{(i)} \neq G^{(j)}$ for $0 \leqq i < j \leqq m$, since therefore all the $L(G^{(i)})$ are different, and since consequently the inequality of the $G \cap L(G^{(i)})$ for $0 \leqq i \leqq m$ may be deduced from II, b.

13.5. *$E(G)$ is a primary ring (with $m(E(G)) = m$)* (I; II, b; IV, 2; V).

PROOF: $E(G)$ contains an identity element 1 and G is an abelian operator group over the ring $E(G)$. It is a consequence of 13.1 that $D(G) = D(G;E(G))$; and it follows from IV, 2 that $D(G;E(G))$ contains at least two independent cycles of maximum order m. We infer from 13.4 and V, c that the criterion **1F** may be applied on the ring $E(G)$ and $E(G)$ is shown to be a primary ring with $m(E(G)) = m$.

We recall that $A(G)$ is the set of all the $E(G)$-automorphisms of the abelian group G which are induced by right-multiplication of G by elements in K; and that $A(G)$ and $K/R(G)$ are isomorphic rings.

13.6. *$R(G) = 0$ so that $A(G)$ and K are essentially the same* (II, b; IV, 1).

PROOF: There exists by II, b one and only one left-annulet containing G. But G is part of both K and $L(R(G))$ so that $K = L(R(G))$ or $0 = R(K) = R(G)$.

13.7. *If the abelian group G is the direct sum of the subgroup U in $D(G)$ and of the cycle $Z = zE(G)$ in $D(G)$, and if the order of the element t in G does not exceed $n(Z)$, then there exists one (and only one) $E(G)$-automorphism in $A(G)$ which maps U upon 0 and z upon t* (I; II, b; III; IV, 2; V; VI).

We note that it is a consequence of 13.1 that the cycles in $D(G)$ are of the form $gE(G)$ for g in G, and that it is a consequence of 13.5 that G (as well as the subgroups in $D(G)$) are primary abelian operator groups over the (primary) ring $E(G)$.

PROOF: There exists by II, b to every S in $D(G)$ one and only one left-annulet S^* in K such that $S = G \cap S^*$, since G is a cycle of maximum order m in $RA(K)$, and since m is by IV, 1 the maximum order of the cycles in $LA(K)$. It is a consequence of 13.1 that a projectivity of $D(G)$ upon $LA(K)$ is effected by mapping S upon S^*.

U is a primary abelian operator group over the (primary) ring $E(G)$. Hence it follows from **1E** that U is the direct sum of cycles Z_v in $D(G)$; and there exist uniquely determined elements c, c_v in Z and Z_v respectively almost all of which are 0, whose sum is t and whose orders do not exceed $n(t) \leqq n(Z)$.

K is the direct sum of Z^* and U^*, since 0 is the cross-cut and K the join of Z^* and U^* in $LA(K)$, since therefore 0 is the cross-cut of $R(Z^*)$ and $R(U^*)$, and since we may apply I. We infer from Lemma 10.1 the existence of idempotents e and f such that $1 = e + f$, $ef = fe = 0$, $Z^* = Ke$, $U^* = Kf$ and $R(U^*) = eK$. Likewise we prove the existence of idempotents e_v such that $Z_v^* = Ke_v$.

Since G is a cycle of maximum order in $RA(K)$, since Z^* is a cycle in $LA(K)$, since K is the direct sum of Z^* and of some left-annulet in K, and since $zK \cap Z^* = zK \cap Ke = zKe \leqq G$ (by Lemma 10.4 and the right-ideal property of G), it follows from III that there exists a right-annulet T in K such that $zKe = T \cap Z^*$. Since the cross-cut of right-annulets is a right-annulet, and since G is a cycle in $RA(K)$, we deduce that $T \cap G = G^{(i)}$ for $0 \leqq i \leqq m$; and we infer from $zKe \leqq G$ that $G^{(i)} \cap Z^* = T \cap G \cap Z^* = zKe \cap G = zKe$. Since z belongs to Z, it belongs to $Z^* = Ke$ so that $z = ze$ belongs to zKe. It is a consequence of 13.1 and 13.2 that $G^{(i)}$ belongs to $D(G)$ and thus $G^{(i)}$ contains with z the subgroup $Z = zE(G)$ in $D(G)$, i.e.

$$Z \leqq G^{(i)} \cap Z^* = zKe \leqq G \cap Z^* = Z \quad \text{or} \quad Z = zKe.$$

Since the element c belongs to Z, and since $z = ze$, there exists an element c' such that $c = zec'e$.

Since U^* is an anti-cycle of order $n(Z)$ in $LA(K)$, it follows that $R(U^*) = eK$ is a cycle of order $n(Z)$ in $RA(K)$; and it follows from VI that $GeK = G^{(m-n(Z))}$ (remembering that G is a cycle of maximum order m in $RA(K)$). It is a consequence of Lemma 10.3 and of 13.3 that $GeKe_v = GeK \cap Ke_v = G^{(m-n)} \cap Z_v^* = G^{(m-n)} \cap G \cap Z_v^* = G^{(m-n)} \cap Z_v = \bar{Z}_v$ where $n = n(Z)$ and where $\bar{Z}_v$ is the uniquely determined subcycle of Z_v in $D(G)$ whose order is the minimum of $n(Z)$ and $n(Z_v)$. Since the order of the element c_v in Z_v does not exceed $n(Z)$, and since $Z = zKe$, it follows that c_v is an element in $\bar{Z}_v = GeKe_v = (G \cap Ke)eKe_v = (G \cap Z^*)eKe_v = ZeKe_v = zKeKe_v$; and hence there exist elements c_v', c_v'' such that $c_v = zec_v'ec_v''e_v$, since $z = ze$. Since almost all the c_v are null, we may assume that almost all the elements $ec_v'ec_v''e_v$ are null; and hence we may form the element $b = ec'e + \sum_v ec_v'ec_v''e_v$. Clearly $eb = b$ so that $Ub \leqq U^*b = Kfeb = 0$; and $zb = zec'e + \sum_v zec_v'ec_v''e_v = c + \sum_v c_v = t$. Thus we have shown that right-multiplication of G by the element b induces an $E(G)$-automorphism in $A(G)$ which maps U upon 0 and z upon t.

13.8. *Every finite*[54] *$E(G)$-automorphism of the abelian group G is contained in* $A(G)$ (I; II, b; III; IV, 2; V; VI).

PROOF: If φ is a finite $E(G)$-automorphism of G, $N(\varphi)$ the set of all the elements in G that are mapped upon 0 by φ, then there exists by Lemma 6.1 a finite number of cycles $z_iE(G) = Z_i$ in $D(G)$ and a subgroup V in $D(G)$ such that G is the direct sum of V and the Z_i and such that $V \leqq N(\varphi)$, since G is by 13.5 a primary abelian operator group over the (primary) ring $E(G)$, and since $D(G) = D(G;E(G))$ by 13.1. Since $n(z_i^{\varphi}) \leqq n(Z_i)$, there exists by 13.7 an $E(G)$-automorphism φ_i in $A(G)$ which maps $V + \sum_{j \neq i} Z_j$ upon 0 and z_i upon z_i^{φ}. It is readily seen that $\varphi = \sum_i \varphi_i$; and consequently φ belongs to $A(G)$, since $A(G)$ is a ring of automorphisms.

13.9. *Every $E(G)$-automorphism of the abelian group G is contained in $A(G)$* (I; II, b; III; IV, 2; V; VI; VII).

---

[54] For definition and properties of finite automorphisms, cp. section 6.

PROOF: If J is a right-annulet, then $J = R(L(J)) \leqq R(L(J) \cap G)$ and $L(J) \cap G \leqq L(R(L(J) \cap G))$. It is a consequence of II, b that $L(J)$ is the only left-annulet in K whose cross-cut with G is just $G \cap L(J)$ and thus we deduce from the last inequality that $L(J) \leqq L(R(L(J) \cap G))$. Hence $R(L(J) \cap G) = R(L(R(L(J) \cap G))) \leqq R(L(J)) = J \leqq R(L(J) \cap G)$ and thus we have shown that every right-annulet J in K satisfies:

$$J = R(L(J) \cap G).$$

To apply postulate VII we have to consider the class W of all the right-annulets in K which are cross-cuts of a finite number of anticycles in $RA(K)$. Clearly the right-annulet J belongs to the class W if, and only if, $L(J)$ is the join of a finite number of cycles in $LA(K)$; and it is a consequence of 13.1 and II, b that the latter property is equivalent to the condition: $L(J) \cap G$ is the sum of a finite number of cycles in $D(G)$.

If φ is an E-automorphism of G, J a right-annulet in the class W, then we denote by $f(J)$ the set of all the elements in K which induce by right-multiplication in G an automorphism ξ in $A(G)$, satisfying: $\xi \equiv \varphi \bmod \mathrm{P}(L(J) \cap G)$, i.e. $f(J)$ consists of all the elements x in K which satisfy: $yx = y^{\varphi}$ for every element y in $L(J) \cap G$.

If J is in the class W, then $L(J) \cap G$ is the sum of a finite number of cycles in $D(G) = D(G;E(G))$ and the right-annulet $\mathrm{P}(L(J) \cap G)$ in the automorphism ring $\mathrm{A}(G;E(G))$ of the primary abelian operator group G over the (primary) ring $E(G)$ (cf. 13.5) is the cross-cut of a finite number of anti-cycles in the partially ordered set of all the right-annulats in $\mathrm{A}(G;E(G))$, as may be inferred from the fundamental theorem of section **2**. Hence it follows from Theorem 6.3, (b) that there exists a finite $E(G)$-automorphism ρ of G satisfying: $\rho \equiv \varphi \bmod \mathrm{P}(L(J) \cap G)$; and this shows by 13.8 that $f(J)$ is not empty.

If the right-annulet H belongs to the class W, and if $H \leqq J$, then $L(J) \cap G \leqq L(H) \cap G$. If furthermore x is an element in $L(J) \cap G$, y an element in $f(J)$ and z an element in $f(H)$, then it follows that $xy = x^{\varphi} = xz$; and this shows that $y - z$ belongs to $R(L(J) \cap G) = J$. Now it is readily seen that $f(J)$ is a coset of K modulo J, and that $H \leqq J$ implies $f(H) \leqq f(J)$, i.e. $f(J)$ is a W-function. Hence there exists by VII one and only one element f in K such that $f(J) = J + f$ for every J in the class W.

If g is any element in G, then there exists by II, b one and only one left-annulet L_g such that $gE(G) = L_g \cap G$. Clearly $J_g = R(L_g)$ belongs to the class W, since $L(J_g) = L(R(L_g)) = L_g$ is a cycle in $LA(K)$. Since g belongs to $L(J_g) \cap G$, and since f belongs to $f(J_g)$, it follows that $gf = g^{\varphi}$; and this shows that φ belongs to $A(G)$.

It is a consequence of 13.5 that G is a primary abelian operator group over the (primary) ring $E(G)$ (with $m(E(G)) = m$); it is a consequence of 13.1 that $D(G;E(G)) = D(G)$ contains at least two independent cycles of maximum order $m = m(E(G))$; it is a consequence of 13.6 that K and $A(G)$ are essentially the same; and it is a consequence of 13.9 that $A(G)$ is exactly the ring of all the $E(G)$-automorphisms of G; and this completes the proof of our theorem.

14. Rings with the same Ideal Theory

The rings K and K^* are said to have *the same ideal theory,* if there exists a transformation f which maps in a biunivoque fashion the set of right-ideals in K upon the set of right-ideals in K^*, the set of left-ideals in K upon the set of left-ideals in K^*, the set of cross-cuts of right- and left-ideals in K upon the set of cross-cuts of right- and left-ideals in K^* and which satisfies the following rules:

(1) If R is a right-ideal in K and L a left-ideal in K, then $(R \cap L)^\varphi = R^\varphi \cap L^\varphi$; and $(LR)^\varphi = L^\varphi R^\varphi$.

(2) If S and T are both right-ideals or both left-ideals in K, then $(ST)^\varphi = S^\varphi T^\varphi$; and S is part of T if, and only if, S^φ is part of T^φ.

THEOREM: *Suppose that G is a primary abelian operator group over the (primary) ring E, that $D(G;E)$ contains at least three independent cycles of maximum order $m(E)$, and that the ring K contains an identity element* 1 *and satisfies postulate VII of section* **12**. *Then the automorphism ring $\Lambda(G;E)$ and the ring K have the same ideal theory if, and only if, they are isomorphic.*

PROOF: Isomorphic rings clearly have the same ideal theory. Thus let us suppose that the rings $\Lambda(G;E)$ and K have the same ideal theory. It is a consequence of the Existence Theorem of section **13** that the postulates I; II, b; III; IV, 3; V; VI are satisfied by the automorphism ring $\Lambda(G;E)$; and consequently the postulates I; II, b; III; IV, 3; V; VI; VII are satisfied by the ring K. Thus we infer from the Existence Theorem of section **13** the existence of a primary abelian operator group G^* over the (primary) ring E^* such that $m(E^*)$ is the maximum order of the cycles in $D(G^*;E^*)$ and such that the rings K and $\Lambda(G^*;E^*)$ are isomorphic. It is a consequence of the Fundamental Theorem of section **2** that there exists a projectivity of $D(G;E)$ upon the partially ordered set $LA(\Lambda(G;E))$ of the left-annulets in $\Lambda(G;E)$ and that there exists a projectivity of $LA(\Lambda(G^*;E^*))$ upon $D(G^*;E^*)$; and there exists a projectivity of $LA(\Lambda(G;E))$ upon $LA(\Lambda(G^*;E^*))$, since $\Lambda(G;E)$ and $\Lambda(G^*;E^*)$ have the same ideal theory. Thus there exists a projectivity of $D(G;E)$ upon $D(G^*;E^*)$; and it follows from **1G** that there exists an isomorphism η of G upon G^* and an isomorphism η of E upon E^* such that $(ge)^\eta = g^\eta e^\eta$ for g in G and e in E. Such an isomorphism η clearly induces an isomorphism of $\Lambda(G;E)$ upon the ring $\Lambda(G^*;E^*)$ isomorphic to K, and this completes the proof.

UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS

BIBLIOGRAPHY

A. A. ALBERT.

(1) *Structure of Algebras.* A. M. S. Colloquium Publications **24**; New York, 1939.

K. ASANO.

(1) *Über verallgemeinerte Abelsche Gruppen mit hyperkomplexem Operatorenring und ihre Anwendungen.* Jap. Journ. of Math. **15** (1939), 231–253.

R. BAER.

(1) *The decomposition of abelian groups into direct summands.* Quart. Journ. of Math. **6** (1935), 222–232.

(2) *The significance of the system of subgroups for the structure of the group.* Amer. Journ. of Math. **61** (1938), 1–44.

(3) *Duality and commutativity of groups.* Duke Math. Journ. **6** (1940), 824–838.
(4) *A unified theory of projective spaces and finite abelian groups.* Transactions Amer. Math. Soc. **52** (1942), 283–343.

R. Brauer.
(1) *Über Systeme hyperkomplexer Zahlen.* Math. Zeitschr. **29** (1929), 79–107.

H. Fitting.
(1) *Die Theorie der Automorphismenringe Abelscher Gruppen und ihr Analogon bei nichtkommutativen Gruppen.* Math. Ann. **107** (1932), 514–542.

G. Köthe.
(1) *Verallgemeinerte Abelsche Gruppen mit hyperkomplexem Operatorenring.* Math. Zeitschr. **39** (1934), 31–44.

P. E. Lewis.
(1) *Characters of abelian groups.* Amer. Journ. of Math. **64** (1942), 81–105.

T. Nakayama.
(1) *On Frobeniusean Algebras.* Ann. of Math. **40** (1939), 611–633; **42** (1941), 1–21.

J. von Neumann.
(1) *Continuous Geometry.* Princeton Lectures 1936/37.

L. Pontrjagin.
(1) *Topological Groups.* Princeton 1939.

M. Shiffman.
(1) *The ring of automorphisms of an abelian group.* Duke Journ. of Math. **5** (1939), 579–597.

K. Shoda.
(1) *Über den Automorphismenring bezw. die Automorphismengruppe einer endlichen Abelschen Gruppe.* Proc. Imp. Acad. Japan, **6** (1930), 10.
(2) *Über die Automorphismen einer endlichen Abelschen Gruppe.* Math. Ann. **100** (1928), 674–686.

Th. Skolem.
(1) *Zur Theorie der assoziativen Zahlensysteme.* Norske Videnskab Selskapet Skrifter Oslo, 1927, (12).

ANNALS OF MATHEMATICS
Vol. 44, No. 2, April, 1943

THEOREMS ON LINEAR COMBINATORIAL TOPOLOGY AND GENERAL MEASURE

BY R. RADO

(Received November 9, 1942)

1. Introduction

1.1 Consider a cube whose vertices are $a_1, a_2, \cdots, a_8$. Give to every one of its 12 edges some arbitrary orientation, denoting by $a_k \rightarrow a_l$ the fact that a_k and a_l are joined by an edge and that, in our orientation, a_k is its first and a_l its second endpoint. We may, for instance, agree to write $a_k \rightarrow a_l$ whenever a_k is nearer to a_1 than a_l.

Associate with every vertex a_n an integer $f(a_n) \geqq 0$. We shall be concerned with certain transformations T which operate on such functions f and which may be described as having the effect of "pushing a unit along some edge $a_k \rightarrow a_l$". More accurately, the transformation $T^{a_k, a_l, 1}$ transforms f into f_1 where

$$\begin{aligned} f_1(a_k) &= f(a_k) - 1 \\ f_1(a_l) &= f(a_l) + 1 \\ f_1(a_n) &= f(a_n) \qquad\qquad (n \neq k, l). \end{aligned}$$

$T^{a_k, a_l, 1}$ is only applicable to f provided $f(a_k) > 0$.

The object of this paper is to investigate, under similar but far more general circumstances, whether it is possible to transform by means of a finite number of suitable T-transformations, a given function $f(a_n)$ into a given function $g(a_n)$. We shall also consider two classes of more specialized T-transformations namely

(i) those (denoted by T') for which $f_1(a_k) \geqq g(a_k)$
(ii) those (denoted by T'') for which $f_1(a_l) \leqq f(a_k)$.

Special cases of the general problem have been treated before (see 1.4).

1.2 Let $\mathfrak{A} = \{a_{n_1}, a_{n_2}, \cdots, a_{n_r}\}$[1] be a non-empty set of vertices which has the following property: *whenever* $a_k \rightarrow a_l \prec \mathfrak{A}$ *then* $a_k \prec \mathfrak{A}$. We call such sets of vertices *sections*. If we adopt, for instance, the convention about $a_k \rightarrow a_l$ mentioned in 1.1, and if $a_1 \rightarrow a_2$, $a_1 \rightarrow a_3$ then $\{a_1, a_2, a_3\}$ is a section. A moment's consideration shows that, for every section $\mathfrak{A}$, the function f_1 of 1.1 satisfies

$$f_1(a_{n_1}) + f_1(a_{n_2}) + \cdots + f_1(a_{n_r}) \leqq f(a_{n_1}) + \cdots + f(a_{n_r}).$$

[1] Brackets { } are only employed to denote sets consisting of the elements specified between them. Relations $X \prec \mathfrak{S} \subset \mathfrak{S}'$ mean that X is element of the set $\mathfrak{S}$ which, in turn, is subset of the set $\mathfrak{S}'$. Sums, differences and products of sets have the meaning usually assigned to them in the theory of point sets.

Hence a *necessary* condition for the possibility of transforming, by means of suitable T-transformations, f into g is that, for every section $\mathfrak{A}$,

$$\sum_{a_n < \mathfrak{A}} f(a_n) \geqq \sum_{a_n < \mathfrak{A}} g(a_n). \tag{1}$$

Our main result states that, under more general circumstances which are explained in 1.3, these conditions are also *sufficient*.

1.3 I shall briefly indicate the kind of generalization to be made. Instead of the system of edges of a cube we shall consider, in the language of combinatorial topology, an arbitrary orientated one-dimensional complex or, using a different terminology, an arbitrary finite oriented graph Γ[2]. The functional values $f(x)$ attached to the "points" x of Γ, will not be non-negative integers but certain sets, subsets of a given abstract set S. The effect of a transformation $T^{a,b,A}$ upon f will be the removal of a subset A from the set $f(a)$ and the addition of the same set A to the set $f(b)$. Here, as before, $a \to b$ is an edge of Γ. Instead of postulating that after a number of suitable T-transformations f is changed into g we only require that f is transformed into a function $\bar{f}$ which is *equimeasurable* to the given function g, a term I am going to explain now.

The sets $f(x)$, $g(x)$ belong to a system of sets $A, B, \cdots$ for which, according to the definitions[3] in **1**, p. 62–64, a monotonic, additive and distributive measure $| A |$ is defined. The relevant definitions and facts are set out in section 2 of this note. All commonly employed types of measure are instances of this general measure definition. The system of sets $\bar{f}(x)$ and the system of sets $g(x)$ are called *equimeasurable* if, from the point of view of the measure in question, they are indistinguishable, i.e.: Whenever $\bar{A}$ is a set obtained from the sets $\bar{f}(x)$ by repeated applications of the processes of addition, subtraction and multiplication, and B is the set obtained by the same processes applied to the corresponding sets $g(x)$ then always $| A | = | B |$. The simple problem of 1.1 is a special case of this more general problem. For suppose that $f(a_n)$, $g(a_n)$ are non-negative integers. Let $f'(a_n)$, $g'(a_n)$ be sets containing exactly $f(a_n)$ and $g(a_n)$ elements respectively, and suppose that for $m \neq n$ the sets $f'(a_m)$, $f'(a_n)$ as well as the sets $g'(a_m)$, $g'(a_n)$ have no element in common. As measure $| A |$ we employ the number of elements of A. Then the generalized problem for the functions f', g' coincides with the special problem for f, g.

Returning to the general case, it will be obvious from fundamental properties of the measure that a *necessary* condition for the possibility of transforming $f(x)$ in the described way into a function $\bar{f}(x)$ which is equimeasurable with $g(x)$ is, that the measure of the set

$$\sum_{x < \mathfrak{A}} f(x)$$

[2] The main result of this note has been extended to the case of an arbitrary *infinite* graph. It is hoped that this case will be dealt with in a later paper.

[3] Numbers in heavy type refer to the bibliography at the end.

is not less than the measure of

$$\sum_{x<\mathfrak{A}} g(x).$$

Here $\mathfrak{A}$ is any section of Γ. It will be proved that these conditions are also *sufficient*, provided the given functions $f(x)$, $g(x)$ are such that sets $f(x)$ belonging to different points x have no element in common, and a similar condition holds for g. The case of arbitrary functions f, g will also be settled. It will be reduced to the special case just described, by familiar methods belonging to Boolean Algebra which have already been used in **1**, p. 79 section 2.

1.4 Theorem 5 below contains as special case the main result of **1** (Theorem I, p. 66). A simple instance where Theorem 4 may be employed is the proof of Muirhead's theorem on inequalities between symmetrical polynomials (**2**, p. 44 ff., in particular p. 46, (2). Also p. 63). Here the graph is

$$a_1 \to a_2 \to \cdots \to a_n\,,$$

and the functional values $f(a_\nu)$ are non-negative numbers.

In a concluding section it is shown that the axioms imposed upon the measure are, in fact, necessary for the validity of our theorems. Furthermore, it is proved that, in the case of any graph and real numbers as functional values of f and g, the conditions (1) are, apart from certain trivial exceptions, independent from each other.

2. Definitions

2.1 Let Γ be a finite orientated graph, i.e. a finite, non-empty set of objects called *points*, here always denoted by the letters a, b, c, x, y (and a' etc.), together with a relation "$\to$" valid between certain ordered pairs of distinct points. $a \to b$ defines an "edge" of Γ.[4] The truth of $a \to b$ does not exclude the truth of $b \to a$, nor does necessarily every point x occur in some relation $a \to x$ or $x \to b$. Indeed, Γ may consist of points only, without any edges joining them. As already mentioned, $x \to x$ does not hold for any x.

Gothic capitals $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$, $\mathfrak{G}$ denote sets of points of Γ. In particular $\mathfrak{G}$ is the set of all points,

$$\mathfrak{G} = \sum_x \{x\}.$$

Unless the contrary is stated, the letter $\mathfrak{A}$ (and $\mathfrak{A}'$ etc.) is used exclusively to denote sets which possess a certain characteristic property and which we call *sections* of Γ. A section is, by definition, a set $\mathfrak{A}$ which is such that *whenever*

$$a \to b < \mathfrak{A}$$

then

$$a < \mathfrak{A}.$$

For instance $\mathfrak{G}$ is a section, and so is the empty set 0.

---

[4] Other terms found in the literature are: node, knot, vertex for point, and branch for edge.

2.2 Let S be an abstract aggregate, fixed throughout. The letters A, B, C denote subsets of S. A measure $| A |$ is defined in S, in the sense of **1,** 62. This means that certain subsets of S are termed *measurable* sets. From now onwards A, B, C always denote measurable sets. We suppose that for every A, B the sets $A + B$; AB; $A - AB$ are measurable. The system of all measurable sets is divided into non-overlapping classes, and the class to which A belongs is denoted by $| A |$ and is called "the measure of A." The system of all measures $| A |$ is ordered by means of a transitive relation "$<$." Given any A, B exactly one of the three relations

$$| A | < | B |; \qquad | A | = | B |; \qquad | A | > | B |$$

holds. $| A | \leqq | B |$ denotes the logical sum of the first two of these relations. The measure is supposed to have the following properties.

Property (M): it is *monotonic*, i.e. $A \subset B$ implies $| A | \leqq | B |$.

Property (A): it is *additive*, i.e.

$$| A_1 | = | B_1 |; \qquad | A_2 | = | B_2 |; \qquad A_1A_2 = B_1B_2 = 0^5$$

imply
$$| A_1 + A_2 | = | B_1 + B_2 |.$$

Property (D): it is *distributive*, i.e.

$$| A | \geqq | B_1 + B_2 |; \qquad B_1B_2 = 0$$

imply the existence of sets A_1, A_2 satisfying

$$A_1 + A_2 \subset A; \qquad A_1A_2 = 0;$$
$$| A_1 | = | B_1 |; \qquad | A_2 | = | B_2 |.$$

In particular it follows from (D) (put $B_2 = 0$) that

$$| A | \geqq | B_1 |$$

implies the existence of $A_1 \subset A$ such that

$$| A_1 | = | B_1 |.$$

For convenience of reference let us call this property of the measure *Property* (C) (the measure is *continuous*).

In fact, (D) is not used explicitly in this paper, everywhere the special case (C) is sufficient. The only occasion where (D) is needed is the proof of Lemma 1, and this proof is not given here since Lemma 1 is taken from **1.** (D) is, however, necessary for Theorem 2 to be true as will be shown in 8.1.

The set A is said to be of *finite measure* (notation: $| A | < \infty$) if $A' \subset A$; $| A' | = | A |$ imply $| A - A' | = 0$.

Instances where this theory of measure applies are:

(i) S is any abstract aggregate, measurable for all subsets of S, and we write $| A | < | B |$ if, and only if, the power of A, as defined in the theory of aggre-

[5] By 0 we denote, without risking any confusion, the number zero as well as the empty set and the measure of the empty set.

gates, is less than the power of B. Here sets of finite measure in our sense are all finite subsets of S.

(ii) S is a euclidean space, its measurable subsets are all Lebesgue-measurable sets in S, and $|A| < |B|$ signifies that the Lebesgue measure of A is less than that of B. Here sets of finite measure in our sense are those of finite Lebesgue measure.

(iii) A more sophisticated example: S is the set of all real numbers, its measurable subsets are all sets A of the form

$$A = \sum_{\nu=1}^{n} [\alpha_\nu, \beta_\nu)$$

where $[\alpha, \beta)$ is the set of all t satisfying $\alpha \leqq t < \beta$, i.e. all sets consisting of a finite number of intervals which are closed on the left and open on the right. Define a number $r(A)$ as follows. If A contains no integer then let $h(A)$ be the (necessarily finite) number of numbers in A which are of the form $k + \frac{1}{l}$ where k and l are integers, $l > 1$. Then put

$$r(A) = \frac{h(A)}{h(A) + 1}.$$

If A contains exactly $m(> 0)$ integers then put

$$r(A) = m.$$

Define a measure by writing $|A| < |B|$ if, and only if $r(A) < r(B)$. It is easily verified that (M), (A), (D) hold. Sets of finite measure are those which contain no integers. If for instance

$$A = [0, 4); \qquad B = [4\tfrac{1}{2}, 5); \qquad C = [5, 6)$$

then

$$0 < |B| < |C| = |B + C| < |A| = |A + B| < |A + C|$$
$$= |A + B + C|; \qquad AB = AC = BC = 0.$$

2.3 Let $f(x)$, $g(x)$ be functions which are defined for every point x, and whose functional values are measurable subsets of S. Throughout this note g will be kept fixed. It is supposed to satisfy

$$(2) \qquad |g(x)| < \infty \qquad \textit{for all } x.$$

g is called *decreasing* if $a \to b$ implies $|g(a)| \geqq |g(b)|$. $f(x)$ is quite arbitrary. It will be subjected to certain transformations T. Choose any edge $a \to b$ and any set $A \subset f(x)$. Then $T^{a,b,A}$ denotes a transformation which transforms $f(x)$ into

$$(3) \qquad T^{a,b,A}f = f_1$$

defined by:

$$f_1(a) = f(a) - A; \qquad f_1(b) = f(b) + A; \qquad f_1(x) = f(x) \quad \textit{for} \quad x \neq a, b.$$

If, in addition, $|f_1(a)| \geqq |g(a)|$ then (3) is called a *T'-transformation.* If $|f_1(b)| \leqq |f(a)|$ we call (3) a *T''-transformation.* A function $\bar{f}$ is called a *T-transform* of f if $\bar{f}$ can be obtained from f by applying a finite number of suitable transformations $T^{a,b,A}$. Analogous meanings are attached to the terms *T'-transform*, *T''-transform*.

Finally, we introduce the abbreviation

$$(f, \mathfrak{B}) = \sum_{x \prec \mathfrak{B}} f(x).^{6}$$

3. Statement of the theorems

3.1 Common hypotheses for theorems 1–4 are (2) and

$$f(x)f(y) = g(x)g(y) = 0 \qquad \text{for } x \neq y. \tag{4}$$

THEOREM 1. *Given $f(x)$ and $g(x)$ necessary and sufficient for f to have a T'-transform $\bar{f}$ satisfying*

$$|\bar{f}(x)| \geqq |g(x)| \qquad \text{for all } x, \tag{5}$$

is that

$$|(f, \mathfrak{A})| \geqq |(g, \mathfrak{A})| \qquad \text{for all } \mathfrak{A}. \tag{6}$$

THEOREM 2. *Given $f(x)$ and $g(x)$, necessary and sufficient for f to have a T'-transform $\bar{f}$ satisfying*

$$|\bar{f}(x)| = |g(x)| \qquad \text{for all } x, \tag{7}$$

is that, in addition to (6),

$$|(f, \mathfrak{A})| = |(g, \mathfrak{A})|. \tag{8}$$

THEOREM 3. *Given $f(x)$ and $g(x)$, and supposing that $g(x)$ is decreasing, necessary and sufficient for f to have a T''-transform $\bar{f}$ satisfying* (5) *is that* (6) *should be true.*

THEOREM 4. *Given $f(x)$ and $g(x)$, and supposing that $g(x)$ is decreasing, necessary and sufficient for f to have a T''-transform $\bar{f}$ satisfying* (7) *is that* (6) *and* (8) *should be true.*

3.2 Before passing to the general case in which (4) is not necessarily satisfied we want to introduce some more notations. A relation

$$a \rightarrow\!\!\rightarrow b$$

means, by definition, that there are points $x_0, x_1, \cdots, x_n$ such that

$$a = x_0 \rightarrow x_1 \rightarrow \cdots \rightarrow x_n = b.$$

Here $n \geqq 0$. In particular, $a \rightarrow\!\!\rightarrow a$ holds for every a, and

$$a \rightarrow\!\!\rightarrow b \rightarrow\!\!\rightarrow c$$

implies $a \rightarrow\!\!\rightarrow c$.

---

[6] Empty sums have the value 0.

Let $[\mathfrak{B}]$[7] denote the smallest section of Γ which contains $\mathfrak{B}$, i.e.

$$[\mathfrak{B}] = \prod_{\mathfrak{B} \subset \mathfrak{A}} \mathfrak{A} = \sum_{x \to \to y \prec \mathfrak{B}} \{x\}.$$

We have, for instance,

$$\mathfrak{A}' = [\mathfrak{A}'] = \sum_{x \prec \mathfrak{A}'} [\{x\}] \tag{9}$$

for every $\mathfrak{A}'$.

If $\mathfrak{B} \neq 0$ we denote by $F(\mathfrak{B})$ the set consisting of those elements of S which belong to every set $f(x)$ when $x \prec \mathfrak{B}$, and to none of the remaining sets $f(x)$, i.e.

$$F(\mathfrak{B}) = \prod_{x \prec \mathfrak{B}} f(x) \prod_{y \text{ not in } \mathfrak{B}} (S - f(y)).^{8} \tag{10}$$

Put $F(0) = 0$. Similarly $\bar{F}(\mathfrak{B})$, $G(\mathfrak{B})$ etc. are defined in terms of $\bar{f}(x)$, $g(x)$ etc. Obviously[9]

$$F(\mathfrak{B}_1)F(\mathfrak{B}_2) = 0 \quad \text{for} \quad \mathfrak{B}_1 \neq \mathfrak{B}_2 .$$

Every set which is obtained from the sets $f(x)$ by the processes of addition, subtraction and multiplication, is a sum of certain sets $F(\mathfrak{B})$. This, in conjunction with property (A), clearly shows that $f(x)$ *and* $g(x)$ *are equimeasurable* (see 1.2) *if, and only if,*

$$| F(\mathfrak{B}) | = | G(\mathfrak{B}) | \qquad \textit{for every } \mathfrak{B}. \tag{11}$$

Let $\psi(x)$ be a notation for any function which associates with every point x a point $\psi(x)$ in such a way that

$$\psi(x) \to\to x \qquad \textit{for all } x.$$

$\psi_{\mathfrak{B}}(x)$ is a notation for those among the $\psi(x)$ which satisfy, in addition,

$$\psi_{\mathfrak{B}}(x_1) \neq \psi_{\mathfrak{B}}(x_2)$$

whenever x_1, x_2 are distinct elements of $\mathfrak{B}$, e.g. $\psi_{\mathfrak{B}}(x) = x$ is such a function. If, in particular, $\mathfrak{B} = 0$ or $\mathfrak{B}$ consists of one single element then every function $\psi(x)$ is at the same time a function $\psi_{\mathfrak{B}}(x)$.

Now put, for $\mathfrak{B} \neq 0$,

$$F^*(\mathfrak{B}) = \Big(\sum_{\psi_{\mathfrak{B}}} \prod_{x \prec \mathfrak{B}} f(\psi_{\mathfrak{B}}(x))\Big) \prod_{y \text{ not in } \mathfrak{B}} (S - f(y)) \tag{12}$$

and

$$F^*(0) = 0.$$

F^*, G^* etc. are defined in terms of $\bar{f}$, g etc. as F^* was defined in terms of f.

In theorems 5 and 6 $f(x)$ and $g(x)$ are arbitrary functions, not necessarily satisfying (4). But (2) is true throughout.

---

[7] Square brackets will only be used in this special sense.

[8] $y \nprec \mathfrak{B}$ is the negation of $y \prec \mathfrak{B}$, empty products have the value S.

[9] A formal proof is given in Footnote [7,2].

THEOREM 5. *Given $f(x)$ and $g(x)$, necessary and sufficient for f to have a T-transform $\bar{f}$ satisfying*

$$(13) \qquad |\bar{F}(\mathfrak{B})| \geqq |G(\mathfrak{B})| \qquad \textit{for all } \mathfrak{B}$$

is that

$$(14) \qquad \left|\sum_{\kappa=1}^{k} F^*(\mathfrak{B}_\kappa)\right| \geqq \left|\sum_{\kappa=1}^{k} G^*(\mathfrak{B}_\kappa)\right|$$

for every $k > 0$ and every selection of k sets $\mathfrak{B}_\kappa$.

THEOREM 6. *Given $f(x)$ and $g(x)$, necessary and sufficient for f to have a T-transform $\bar{f}$ which is equimeasurable to g, i.e. which satisfies*

$$(15) \qquad |\bar{F}(\mathfrak{B})| = |G(\mathfrak{B})| \qquad \textit{for all } \mathfrak{B}$$

is that (8) *and* (14) *should be true.*

Evidently (15) and (7) are equivalent statements if (4) holds.

3.3 Let us consider some special cases of theorems 5 and 6. The necessary verifications are easily made and are left to the reader.

(i) $\mathfrak{G} = \{1, 2, \cdots, n, n+1\}$. The edges of Γ are

$$\nu \to n+1 \qquad (1 \leqq \nu \leqq n).$$

If $\mathfrak{B} \subset \{1, 2, \cdots, n\}$ then

$$F^*(\mathfrak{B}) = F(\mathfrak{B}),$$
$$F^*(\mathfrak{B} + \{n+1\}) = \Big(\prod_{x \prec \mathfrak{B}} f(x)\Big)\Big(\sum_{y \text{ not in } \mathfrak{B}} f(y)\Big).$$

Now suppose that, in particular,

$$f(n+1) = \sum_{\nu=1}^{n} f(\nu); \qquad g(n+1) = \sum_{\nu=1}^{n} g(\nu),$$

while

$$f(1), \cdots, f(n), \qquad g(1), \cdots, g(n)$$

are $2n$ arbitrary sets the $g(\nu)$ $(1 \leqq \nu \leqq n)$, however, of finite measure[10]. Then the only effect of a transformation T upon the function $f(x)$ is the removal of some set A from one of the sets $f(\nu)$ $(1 \leqq \nu \leqq n)$, and vice versa, every such removal is effected by a transformation T. Furthermore, for every $\mathfrak{B} \subset \{1, \cdots, n\}$, we have

$$F^*(\mathfrak{B}) = 0,$$
$$F^*(\mathfrak{B} + \{n+1\}) = \prod_{x \prec \mathfrak{B}} f(x) \qquad (\textit{if } \mathfrak{B} \neq 0),$$
$$F^*(\{n+1\}) = f(n+1),$$

---

[10] By Lemma 1 below, $g(1) + \cdots + g(n)$ is of finite measure.

and analogous equations with g instead of f. Hence Theorem 6 reduces to the main result of **1** (Theorem I) which states:

Given $2n$ *sets* $A_1, \cdots, A_n, B_1, \cdots, B_n$, *of which the* B_ν *are of finite measure, it is possible to select subsets* $\bar{A}_\nu \subset A_\nu$ *in such a way that the system* $\bar{A}_1, \cdots, \bar{A}_n$ *is equimeasurable to the system* $B_1, \cdots, B_n$, *provided the following (necessary and sufficient) condition holds: For every set* A *obtained from the* A_ν *by any applications of the processes of addition and multiplication, and for the set* B *obtained by applying the same processes to the corresponding sets* B_ν, *we have* $|A| \geqq |B|$.

(ii) $\mathfrak{G} = \{1, 2, \cdots, n\}$. The edges of Γ are

$$1 \to 2 \to 3 \to \cdots \to n \to 1.$$

If $k > 0; 1 \leqq \nu_1 < \nu_2 < \cdots < \nu_k \leqq n$ then

$$F^*(\{\nu_1, \cdots, \nu_k\}) = \sum_{1 \leqq \lambda_1 < \cdots < \lambda_k \leqq n} f(\lambda_1)f(\lambda_2) \cdots f(\lambda_k).$$

Hence the result: *Given sets* $A_1, \cdots, A_n, B_1, \cdots, B_n$ $(|B_\nu| < \infty)$ *it is possible, by removing suitable subsets from some* A_ν*'s and adding them to the corresponding set* $A_{\nu+1}$ $(A_{n+1} = A_1)$ *to change* $A_1, \cdots, A_n$ *into a system equimeasurable to* $B_1, \cdots, B_n$ *provided the following (necessary and sufficient) conditions hold:*

$$\begin{cases} |A_1 + A_2 + \cdots + A_n| = |B_1 + \cdots + B_n| \\ |A_1A_2 + A_1A_3 + \cdots + A_{n-1}A_n| \geqq |B_1B_2 + \cdots + B_{n-1}B_n| \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ |A_1A_2 \cdots A_n| \geqq |B_1B_2 \cdots B_n|. \end{cases}$$

The kth relation (16) involves all $\binom{n}{k}$ products of k sets A_ν and B_ν respectively.

In fact, (16) is the solubility-condition for any graph with n points which satisfies $x \to\to y$ for every x, y.

(iii) Γ is arbitrary, f and g satisfy (4). Then

$$F(\{a\}) = f(a),$$

$$F^*(\{a\}) = \sum_{x \to\to a} f(x) = (f, [\{a\}]),$$

$$F(\mathfrak{B}) = F^*(\mathfrak{B}) = 0$$

if $\mathfrak{B}$ contains more than one element, (14) becomes

$$\left| \sum_{\kappa=1}^{k} (f, [\{a_\kappa\}]) \right| \geqq \left| \sum_{\kappa=1}^{k} (g, [\{a_\kappa\}]) \right|,$$

where $k > 0$ and the a_κ are arbitrary. By (9) this is the same as

$$|(f, \mathfrak{A})| \geqq |(g, \mathfrak{A})|$$

where

$$\mathfrak{A} = \sum_{\kappa=1}^{k} [\{a_\kappa\}]$$

is any arbitrary section. Hence Theorem 5 reduces to Theorem 1 if we weaken the latter's assertion by writing T for T'.

4. Preliminary remarks on theorems 1–4

4.1 Property (A) of the measure shows that, whenever

$$|A_\nu| = |B_\nu| \qquad (1 \leqq \nu \leqq 3),$$
$$A_\mu A_\nu = B_\mu B_\nu = 0 \qquad (1 \leqq \mu < \nu \leqq 3)$$

we have

$$|A_1 + A_2| = |B_1 + B_2|,$$
$$|A_1 + A_2 + A_3| = |(A_1 + A_2) + A_3| = |(B_1 + B_2) + B_3| = |B_1 + B_2 + B_3|.$$

Similarly for sums of more than three terms.

All facts on sets of finite measure which we shall have to use are contained in

LEMMA 1. (i) *If* $|A| \leqq |B| < \infty$ *then* $|A| < \infty$.

(ii) *If* $|A| < \infty$; $|B| < \infty$, *then* $|A + B| < \infty$.

(i) is Lemma 5, (ii) is Lemma 7 of **1**.

4.2 LEMMA 2. *If* $n \geqq 2$,

$$|A_\nu| \leqq |B_\nu| \qquad (1 \leqq \nu \leqq n), \tag{17}$$
$$A_\mu A_\nu = B_\mu B_\nu = 0 \qquad (1 \leqq \mu < \nu \leqq n)$$

then

$$|A_1 + \cdots + A_n| \leqq |B_1 + \cdots + B_n|. \tag{18}$$

If $|A_\nu| < \infty$ *for all* ν, *and if there is equality in* (18) *then there is equality in* (17) *for every* ν.

PROOF. By (17) and (C) there are sets $B'_\nu \subset B_\nu$ such that

$$|A_\nu| = |B'_\nu| \qquad (1 \leqq \nu \leqq n).$$

Then, by (A), in the form 4.1, and (M),

$$\left|\sum_1^n A_\nu\right| = \left|\sum_1^n B'_\nu\right| \leqq \left|B_1 + \sum_2^n B'_\nu\right| \leqq \left|\sum_1^n B_\nu\right|, \tag{19}$$

and therefore (18). If $|A_\nu| < \infty$ for all ν, and if there is equality in (18) then, by (19) and Lemma 1,

$$\sum_1^n B'_\nu \subset B_1 + \sum_2^n B'_\nu$$
$$\left|\sum_1^n B'_\nu\right| = \left|B_1 + \sum_2^n B'_\nu\right| \leqq \left|\sum_1^n B_\nu\right| = \left|\sum_1^n A_\nu\right| < \infty,$$
$$|B_1 - B'_1| = 0,$$
$$|A_1| = |B'_1| = |B'_1 + 0| = |B'_1 + (B_1 - B'_1)| \qquad (by\ (A))$$
$$= |B_1|.$$

Therefore

$$|A_\nu| = |B_\nu| \qquad (1 \leqq \nu \leqq n),$$

and the lemma is proved.

4.3 If $a \rightarrow b$, and if $\mathfrak{A}$ is any section the

either $$a, b \prec \mathfrak{A}$$

or $$a \prec \mathfrak{A}; \quad b \nprec \mathfrak{A}$$

or $$a, b \nprec \mathfrak{A}.$$

For the remaining possibility

$$a \nprec \mathfrak{A}; \quad b \prec \mathfrak{A}$$

is excluded by the definition of sections. Now let

$$T^{a,b,A}f = f_1.$$

Then it follows that

$$(f_1, \mathfrak{A}) = \begin{cases} (f, \mathfrak{A}) - A & \text{if } a \prec \mathfrak{A}; b \nprec \mathfrak{A} \\ (f, \mathfrak{A}) & \text{otherwise.} \end{cases}$$

Hence every T-transform $\bar{f}$ of f satisfies

$$(\bar{f}, \mathfrak{A}) \subset (f, \mathfrak{A}).$$

If, in addition, (5) holds then, by (M) and Lemma 2,

$$|(f, \mathfrak{A})| \geqq |(\bar{f}, \mathfrak{A})| \geqq |(g, \mathfrak{A})|.$$

Therefore (6) is a necessary condition in theorems 1–4.

Clearly, the same function $\bar{f}$ satisfies

$$(\bar{f}, \mathfrak{A}) = (f, \mathfrak{A}).$$

If, moreover, (7) is satisfied then

$$|(f, \mathfrak{A})| = |(\bar{f}, \mathfrak{A})| = |(g, \mathfrak{A})|, \tag{20}$$

i.e. (8). We have thus shown that (6) and (8) are necessary conditions in theorems 2 and 4.

4.4 Theorems 2 and 4 are easily deduced from theorems 1 and 3 respectively. For suppose that (6) and (8) are satisfied, and that, (in the case of Theorem 4) $g(x)$ is decreasing. Then by Theorem 1 (or Theorem 3) there is a T'-transform (or a T''-transform) $\bar{f}$ of f for which (5) holds. Then, once more, (20) follows, and (7) is a consequence of Lemma 2. Here we must observe that it follows from (2), (20) and Lemma 1 that $|\bar{f}(x)| < \infty$.

To sum up, all we have to show in order to complete the proofs of theorems 1–4 is that (6) is sufficient for the existence of a T'-transform and, in case g is decreasing, of a T''-transform $\bar{f}$ satisfying (5).

5. Proof of theorems 1 and 2

5.1 Throughout this section (2) and (4) are supposed to hold. We introduce some further definitions and notations. A relation

$$a \equiv b \tag{21}$$

is, by definition, equivalent to

$$a \rightarrow\rightarrow b \rightarrow\rightarrow a,$$

and $a \not\equiv b$ is the negation of (21). An edge $a \rightarrow b$ is called *singular* if $a \equiv b$, and *regular* if $a \not\equiv b$. Singular edges are those which occur in closed cycles

$$x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow \cdots \rightarrow x_n \rightarrow x_1 .$$

Let $\rho(\Gamma)$ denote the number of regular edges of Γ. (21) defines an equivalence relation and hence a subdivision of $\mathfrak{A}$ into mutually exclusive classes. The class to which any point x_o belongs is denoted by

$$\mathfrak{K}(x_o) = \sum_{x \equiv x_0} \{x\}.$$

Finally, put

$$p(x_o, \Gamma) = \sum_{x \rightarrow\rightarrow x_0} 1,$$

$$q(f, x_o, \Gamma) = \sum_{\substack{x \\ |f(x)| \leqq |f(x_0)|}} 1,$$

$$s(f, \Gamma) = \sum_x p(x, \Gamma) q(f, x, \Gamma).$$

p counts the number of "ancestors" of x_o while q describes the position of $|f(x_o)|$ relative to all measures $|f(x)|$.

5.2 Lemma 3. *If* $m \geqq 1$; $|A_1| \leqq |B_1|$,

$$|A_\mu + A| > |B_\mu| \qquad (1 \leqq \mu \leqq m) \tag{22}$$

then there is $A' \subset A$ *and* μ_o $(1 \leqq \mu_o \leqq m)$ *such that*

$$|A_\mu + A'| \begin{cases} \geqq |B_\mu| & (1 \leqq \mu \leqq m) \\ = |B_\mu| & (\mu = \mu_o). \end{cases}$$

This is Lemma 2 of **1**. Its assertion remains, of course, valid if in (22) we allow "$\geqq$" instead of "$>$."

Lemma 4. *Let*

$$a_o \rightarrow b_o ,$$

$$a_o \prec \mathfrak{A}_o ; \qquad b_o \nprec \mathfrak{A}_o .$$

Suppose Γ' *is the graph which is obtained from* Γ *by removing the edge* $a_o \rightarrow b_o$, *i.e.*

(i) $\mathfrak{A}' = \mathfrak{A}$;

(ii) *$x \to y$ is true in Γ' if, and only if, $x \to y$ is true in Γ and $(x, y) \neq (a_o, b_o)$. Let $\mathfrak{A}'$ be a section of Γ' (but not necessarily a section of Γ). Then $\mathfrak{A}_o + \mathfrak{A}'$ and $\mathfrak{A}_o\mathfrak{A}'$ are sections of Γ.*

PROOF. In the following proof the symbol "$\to$" is only used to denote edges of Γ. The characteristic property of $\mathfrak{A}'$ is: *Whenever*

$$a \to b \prec \mathfrak{A}'; \qquad (a, b) \neq (a_o, b_o)$$

then

$$a \prec \mathfrak{A}'.$$

1) If $\mathfrak{A}_o + \mathfrak{A}'$ would not be a section of Γ then there would exist points a, b such that

$$a \to b \prec \mathfrak{A}_o + \mathfrak{A}'; \qquad a \not\prec \mathfrak{A}_o + \mathfrak{A}'.$$

Then we would conclude that

$$a \not\prec \mathfrak{A}_o; \qquad b \not\prec \mathfrak{A}_o; \qquad b \prec \mathfrak{A}'; \qquad a \neq a_o; \qquad a \prec \mathfrak{A}'$$

which is a contradiction.

2) If $\mathfrak{A}_o\mathfrak{A}'$ would not be a section of Γ then there would exist points a, b such that

$$a \to b \prec \mathfrak{A}_o\mathfrak{A}'; \qquad a \not\prec \mathfrak{A}_o\mathfrak{A}'.$$

Then we would conclude that

$$b \prec \mathfrak{A}_o; \qquad a \prec \mathfrak{A}_o; \qquad a \not\prec \mathfrak{A}'; \qquad b \neq b_o; \qquad b \not\prec \mathfrak{A}'$$

which is a contradiction.

5.3·LEMMA 5. *Notations and hypotheses as in Lemma 4: In addition we suppose that* (6) *holds and that*

$$|(f, \mathfrak{A}_o)| = |(g, \mathfrak{A}_o)|.$$

Then

$$|(f, \mathfrak{A}')| \geqq |(g, \mathfrak{A}')| \tag{23}$$

for all sections $\mathfrak{A}'$ of Γ'.

PROOF.

$$\mathfrak{A}' = \mathfrak{A}_o\mathfrak{A}' + ((\mathfrak{A}_o + \mathfrak{A}') - \mathfrak{A}_o),$$

$$(f, \mathfrak{A}') = (f, \mathfrak{A}_o\mathfrak{A}') + (f, (\mathfrak{A}_o + \mathfrak{A}') - \mathfrak{A}_o). \tag{24}$$

By Lemma 4 and (6),

$$|(f, \mathfrak{A}_o\mathfrak{A}')| \geqq |(g, \mathfrak{A}_o\mathfrak{A}')|, \tag{25}$$

$$|(f, \mathfrak{A}_o + \mathfrak{A}')| \geqq |(g, \mathfrak{A}_o + \mathfrak{A}')|. \tag{26}$$

If we assume that

$$(27) \qquad |(f, (\mathfrak{A}_o + \mathfrak{A}') - \mathfrak{A}_o)| < |(g, (\mathfrak{A}_o + \mathfrak{A}') - \mathfrak{A}_o)|$$

we obtain a contradiction. For then, by (2) and Lemma 1,

$$|(f, (\mathfrak{A}_o + \mathfrak{A}') - \mathfrak{A}_o)| < \infty,$$
$$|(f, \mathfrak{A}_o)| = |(g, \mathfrak{A}_o)| < \infty,$$

and hence, by Lemma 2,

$$|(f, \mathfrak{A}_o + \mathfrak{A}')| = |(f, (\mathfrak{A}_o + \mathfrak{A}') - \mathfrak{A}_o) + (f, \mathfrak{A}_o)|$$
$$< |(g, (\mathfrak{A}_o + \mathfrak{A}') - \mathfrak{A}_o) + (g, \mathfrak{A}_o)| = |(g, \mathfrak{A}_o + \mathfrak{A}')|$$

which contradicts (26). Hence (27) is not true, i.e.

$$(28) \qquad |(f, (\mathfrak{A}_o + \mathfrak{A}') - \mathfrak{A}_o)| \geqq |(g, (\mathfrak{A}_o + \mathfrak{A}') - \mathfrak{A}_o)|.$$

(23) follows from (24), (25), (28) and Lemma 2.

5.4 Lemma 6. *Let* $\mathfrak{K}_o = \mathfrak{K}(x_o)$;

$$(29) \qquad |(f, \mathfrak{K}_o)| \geqq |(g, \mathfrak{K}_o)|.$$

Then there exists a T'-transform $\bar{f}$ of f satisfying

$$|\bar{f}(x)| \geqq |g(x)| \qquad (x \prec \mathfrak{K}_o),$$
$$\bar{f}(x) = f(x) \qquad (x \not\prec \mathfrak{K}_o).$$

Proof. Define an operator O_1 as follows. O_1 is applicable to every function $f(x)$ which satisfies (29) and, in addition,

$$(30) \qquad |f(x_1)| < |g(x_1)|$$

for at least one $x_1 \prec \mathfrak{K}_o$. We may assume that O_1 is applicable to the given function f. Choose one point $x_1 \prec \mathfrak{K}_o$ which satisfies (30).[11] Then there is $x_2 \prec \mathfrak{K}_o$ such that

$$|f(x_2)| > |g(x_2)|.$$

For if we assume that

$$|f(x)| \leqq |g(x)| \quad \textit{for every} \quad x \prec \mathfrak{K}_o$$

then Lemma 2, in connection with (30), would lead to a contradiction against (29). It now follows from the definition of $\mathfrak{K}_o$ that there exists a chain of edges

$$x_2 = a_o \to a_1 \to \cdots \to a_n = x_1.$$

[11] We shall often have to define operations which necessitate a uniformly bounded number of selections of elements from sets whose elements are points or edges of sections. We make these operations definite by numbering, once for all, the points, edges and sections of Γ and choosing every time the first one. Selections made in virtue of (D) cannot in general be made definite.

Then $n > 0$. By definition of $\mathfrak{K}_o$ we have $x_1 \to\!\!\to x_2$ and therefore

$$a_\nu < \mathfrak{K}_o \qquad (0 \leqq \nu \leqq n).$$

There is a largest index ν_o satisfying

$$0 \leqq \nu_o < n; \qquad | f(a_{\nu_o}) | > | g(a_{\nu_o}) |,$$

and then a smallest index ν_1 satisfying

$$\nu_o < \nu_1 \leqq n; \qquad | f(a_{\nu_1}) | \neq | g(a_{\nu_1}) |.$$

Then

$$| f(a_\nu) | \begin{cases} > | g(a_\nu) | & (\nu = \nu_o) \\ = | g(a_\nu) | & (\nu_o < \nu < \nu_1) \\ < | g(a_\nu) | & (\nu = \nu_1). \end{cases}$$

(i) If $\nu_o + 1 < \nu_1$ then there exists, by (C), a set $A_o \subset f(a_{\nu_o})$ such that

$$| f(a_{\nu_o}) - A_o | = | g(a_{\nu_o}) |.$$

Then

$$T^{a_{\nu_o}, a_{\nu_o+1}, A_o} f = f_1,$$

is a T'-transform of f,

$$| f_1(a_{\nu_o}) | = | g(a_{\nu_o}) |,$$
$$| f_1(a_{\nu_o+1}) | = | f(a_{\nu_o+1}) + A_o | \geqq | f(a_{\nu_o+1}) |.$$

In the last relation equality is impossible. For if

$$| f(a_{\nu_o+1}) + A_o | = | f(a_{\nu_o+1}) | = | g(a_{\nu_o+1}) | < \infty$$

then, in view of

$$A_o f(a_{\nu_o+1}) \subset f(a_\nu) f(a_{\nu_o+1}) = 0,$$

we would conclude that $| A_o | = 0$ and therefore

$$| g(a_{\nu_o}) | = | f(a_{\nu_o}) - A_o | = | f(a_{\nu_o}) |$$

which is a contradiction. Hence

$$| f_1(a_\nu) | \begin{cases} > | g(a_\nu) | & (\nu = \nu_o + 1) \\ = | g(a_\nu) | & (\nu_o + 1 < \nu < \nu_1) \\ < | g(a_\nu) | & (\nu = \nu_1). \end{cases}$$

(ii) If $\nu_o + 2 < \nu_1$ then we treat f_1, $\nu_o + 1$ in the same way as f, ν_o were treated in (i), etc. After $\nu_1 - \nu_o - 1$ such T'-transformations (no transformation at all if $\nu_o + 1 = \nu_1$) we arrive at a T'-transform f' of f which satisfies

$$| f'(a_{\nu_1-1}) | > | g(a_{\nu_1-1}) |,$$
$$| f'(a_{\nu_1}) | < | g(a_{\nu_1}) |,$$
$$| f'(x) | = | f(x) | \qquad \textit{for } x \neq a_{\nu_o}, a_{\nu_1-1}, a_{\nu_1}.$$

Also,

$$|f'(a_{\nu_o})| = |g(a_{\nu_o})| \qquad \textit{if } \nu_o + 1 < \nu_1 .$$

There is a set $A' \subset f'(a_{\nu_1-1})$ satisfying

$$|f'(a_{\nu_1-1}) - A'| = |g(a_{\nu_1-1})| .$$

Then

$$T^{a_{\nu_1-1}, a_{\nu_1}, A'} f' = f''$$

is a T'-transformation. We put, by definition,

$$O_1 f = f''$$

Then

$$|f''(a_\nu)| = |g(a_\nu)| \qquad (\nu_o \leqq \nu < \nu_1),$$

$$f''(x) = f(x) \qquad (x \neq a_{\nu_o}, a_{\nu_o+1}, \cdots, a_{\nu_1}).$$

In particular, there are more points x satisfying

$$|f''(x)| = |g(x)|$$

than points x satisfying

$$|f(x)| = |g(x)| .$$

For every x of the second kind is also of the first kind, and

$$|f''(a_{\nu_o})| = |g(a_{\nu_o})| ; \qquad |f(a_{\nu_o})| > |g(a_{\nu_o})| .$$

Therefore O_1 can only be applied to f a bounded number of times in succession, and the result of the last possible application is a T'-transform $\bar{f}$ of f which has the required properties.

5.5 LEMMA 7. *Let $\mathfrak{K}_o = \mathfrak{K}(x_o)$. Suppose that*

(31) $$|f(x)| \geqq |g(x)| \qquad (x \prec \mathfrak{K}_o).$$

Then there is a T'-transform $\bar{f}$ of f satisfying

$$|\bar{f}(x)| = |g(x)| \qquad (x \prec \mathfrak{K}_o - \{x_o\}),$$

$$\bar{f}(x) = f(x) \qquad (x \nprec \mathfrak{K}_o).$$

PROOF. Define an operator O_2 as follows. O_2 is applicable to every f which satisfies (31) and, in addition,

$$|f(x_1)| > |g(x_1)|$$

for at least one point $x_1 \prec \mathfrak{K}_o - \{x_o\}$. We may assume that O_2 is applicable to the given function f. Choose one of the above-mentioned points x_1. Then there is a chain

$$x_1 = a_o \to a_1 \to \cdots \to a_n = x_o ,$$

where $n > 0$ and all a_ν are in $\mathfrak{K}_o$. There is a set $A_o \subset f(a_o)$ such that

$$| f(a_o) - A_o | = | g(a_o) | .$$

Then

$$| f(a_1) + A_o | > | g(a_1) | .$$

For otherwise we would have

$$| g(a_1) | \leqq | f(a_1) | \leqq | f(a_1) + A_o | \leqq | g(a_1) | < \infty ,$$

$$| A_o | = 0,$$

$$| g(a_o) | = | f(a_o) - A_o | = | f(a_o) |$$

which, in fact, is not true. Put

$$T^{a_o, a_1, A_o} f = f_1 .$$

This is a T'-transformation. f_1 satisfies

$$| f_1(a_o) | = | g(a_o) |,$$

$$| f_1(a_1) | = | f(a_1) + A_o | > | g(a_1) |,$$

$$f_1(x) = f(x) \qquad (x \neq a_o , a_1).$$

Now treat f_1, a_1 in the same way as f, a_o were treated, etc. until, after n such T'-transformations, we obtain a function f_n satisfying

$$| f_n(x) | \begin{cases} = | g(x) | & (x = a_o , a_1 , \cdots , a_{n-1}) \\ > | g(x) | & (x = a_n), \end{cases}$$

$$f_n(x) = f(x) \qquad (x \neq a_o , \cdots , a_n).$$

Put

$$O_2 f = f_n .$$

The number of points x for which

$$| f_n(x) | = | g(x) |$$

is greater than or equal to the number for which

$$| f(x) | = | g(x) | .$$

For every x of the second kind is also of the first kind, and

$$| f_n(a_o) | = | g(a_o) | ; \qquad | f(a_o) | > | g(a_o) | .$$

Clearly then, in order that the number of points of the second kind be equal to the number of the first kind, it is necessary that $| f(a_n) | = | g(a_n) |$, and this can be true only prior to the first application of O_2.

Hence O_2 can only be applied to f a bounded number of times, and after the

last possible application a function $\bar{f}$ is obtained which has the required properties.

5.6 We now come to the proof of Theorem 1. We assume that (2), (4) and (6) hold, and we have to show the existence of a T'-transform $\bar{f}$ of f which satisfies (5).

CASE 1. Suppose that $\rho(\Gamma) = 0$. Then we define an operator O_3 as follows. O_3 is applicable to every f which satisfies (6) and, moreover,

$$|f(x_o)| < |g(x_o)|.$$

for at least one point x_o. We may assume that O_3 is applicable to the given function f. Choose one such point x_o. The set

$$\mathfrak{K}_o = \mathfrak{K}(x_o)$$

is a section of Γ. For if $x \to y < \mathfrak{K}_o$ then, in view of $\rho(\Gamma) = 0$, we have

$$y \equiv x; \qquad x < \mathfrak{K}_o .$$

(6) implies (29). Hence there exists, by Lemma 6, a T'-transform f' of f which satisfies

$$|f'(x)| \begin{cases} \geqq |g(x)| & (x < \mathfrak{K}_o) \\ = |f(x)| & (x \not< \mathfrak{K}_o). \end{cases}$$

Put

$$O_3 f = f'.$$

Since the number of points x satisfying

$$|f'(x)| \geqq |g(x)|$$

exceeds the number of points x satisfying

$$|f(x)| \geqq |g(x)|$$

it follows that O_3 can only be applied a bounded number of times to the given function f, and after the last possible application we obtain a function $\bar{f}$ which satisfies (5).

CASE 2. Suppose that $\rho(\Gamma) > 0$. We use induction with respect to $\rho(\Gamma)$. Choose a regular edge $a \to b$ for which $p(a)$ reaches its smallest possible value in Γ. Then

$$\mathfrak{K}_o = \mathfrak{K}(a)$$

is a section. For if there would be any points x, y satisfying

$$x \to y < \mathfrak{K}_o ; \qquad x \not< \mathfrak{K}_o$$

then

$$x \not\equiv y; \qquad p(x) < p(y) = p(a)$$

which is impossible. Once more, because of (6), (29) is true, and therefore there exists a T'-transform f' of f satisfying

$$|f'(x)| \geqq |g(x)| \qquad (x \prec \mathfrak{K}_o),$$

$$f'(x) = f(x) \qquad (x \nprec \mathfrak{K}_o). \tag{32}$$

Now Lemma 7 shows the existence of a T'-transform f'' of f' which satisfies

$$|f''(x)| = |g(x)| \qquad (x \prec \mathfrak{K}_o - \{a\}), \tag{33}$$

$$f''(x) = f'(x) \qquad (x \nprec \mathfrak{K}_o). \tag{34}$$

Then

$$|f''(a)| \geqq |g(a)|. \tag{35}$$

For let us suppose that

$$|f''(a)| < |g(a)| < \infty. \tag{36}$$

Then Lemma 2, applied to (33), (36), would lead to

$$|(f'', \mathfrak{K}_o)| < |(g, \mathfrak{K}_o)| \tag{37}$$

while, on the other hand, by (32) and (34),

$$|(f'', \mathfrak{K}_o)| = |(f'', \mathfrak{A}) - (f'', \mathfrak{A} - \mathfrak{K}_o)| = |(f, \mathfrak{A}) - (f, \mathfrak{A} - \mathfrak{K}_o)| = |(f, \mathfrak{K}_o)|.$$

This, together with (29), contradicts (37). Therefore (35). Any section $\mathfrak{A}$ which contains at least one point of $\mathfrak{K}_o$ contains the whole set $\mathfrak{K}_o$. For if

$$x \prec \mathfrak{A}\mathfrak{K}_o; \qquad y \prec \mathfrak{K}_o$$

then

$$y \longrightarrow\!\!\longrightarrow x \prec \mathfrak{A}; \qquad y \prec \mathfrak{A}.$$

Therefore, for every $\mathfrak{A}$,

$$(f'', \mathfrak{A}) = (f, \mathfrak{A}),$$

$$|(f'', \mathfrak{A})| = |(f, \mathfrak{A})| \geqq |(g, \mathfrak{A})|. \tag{37a}$$

By (35) and (C), we can choose a set $A \subset f''(a)$ such that

$$|f''(a) - A| = |g(a)|. \tag{38}$$

Denote by

$$\mathfrak{A}_1, \mathfrak{A}_2, \cdots, \mathfrak{A}_m$$

all sections which have the property that

$$a \prec \mathfrak{A}_\mu; \qquad b \nprec \mathfrak{A}_\mu.$$

$\mathfrak{K}_o$ is one of the sections $\mathfrak{A}_\mu$. For if b were in $\mathfrak{K}_o$ then we would have $a \equiv b$ which contradicts the fact that $a \to b$ is a regular edge. Let

$$\mathfrak{A}_1 = \mathfrak{K}_o .$$

Put

$$(f'', \mathfrak{A}_\mu) = A_\mu ; \qquad (g, \mathfrak{A}_\mu) = B_\mu \qquad (1 \leqq \mu \leqq m).$$

Then, by (33) and (38),

$$| A_1 = A | = | (f'', \mathfrak{K}_o - \{a\}) + (f''(a) - A) | = | (g, \mathfrak{K}_o - \{a\}) + g(a) |$$
$$= | (g, \mathfrak{K}_o) | = | B_1 |.$$

Also, from (37a),

$$| (A_\mu - A) + A | = | A_\mu | \geqq | B_\mu | \qquad (1 \leqq \mu \leqq m).$$

Hence, by Lemma 3, with $A_\mu - A$ instead of A_μ, there exists a set $A' \subset A$ and an index $\mu_o(1 \leqq \mu_o \leqq m)$ such that

$$\begin{array}{ll} (39) & \\ & | (A_\mu - A) + A' | \begin{cases} \geqq | B_\mu | \, (1 \leqq \mu \leqq m) \\ = | B_\mu | \, (\mu = \mu_o). \end{cases} \\ (40) & \end{array}$$

Then

$$T^{a,b,A-A'}f'' = f'''$$

is a T'-transformation. For

$$| f''(a) - (A - A') | \geqq | f''(a) - A | = | g(a) |.$$

Also, by (39),

$$| (f''', \mathfrak{A}_\mu) | \geqq | (g, \mathfrak{A}_\mu) | \qquad (1 \leqq \mu \leqq m).$$

If $\mathfrak{A}$ is a section which does not occur among the sections $\mathfrak{A}_\mu$ then

$$\textit{either} \quad a, b \prec \mathfrak{A}$$
$$\textit{or} \quad a, b \nprec \mathfrak{A}.$$

In both cases

$$(f''', \mathfrak{A}) = (f'', \mathfrak{A}).$$

Hence for every section $\mathfrak{A}$

$$| (f''', \mathfrak{A}) | \geqq | (g, \mathfrak{A}) |.$$

In particular, by (40),

$$| (f''', \mathfrak{A}_{\mu_o}) | = | A_{\mu_o} - (A - A') | = | B_{\mu_o} | = | (g, \mathfrak{A}_{\mu_o}) |.$$

Therefore Lemma 5 is applicable, with

$$\mathfrak{A}_o = \mathfrak{A}_{\mu_o}; \qquad a_o = a; \qquad b_o = b.$$

Using the notation of Lemma 5 we conclude that

$$|(f''', \mathfrak{A}')| \geqq |(g, \mathfrak{A}')|$$

for every section $\mathfrak{A}'$ of Γ'. Now

$$\rho(\Gamma') = \rho(\Gamma) - 1.$$

For if $x \rightarrow y$ is a singular edge of Γ then there exists a closed cycle of edges of Γ:

$$x = x_o \rightarrow y = x_1 \rightarrow x_2 \rightarrow \cdots \rightarrow x_n = x. \tag{41}$$

It follows that every edge $x_{\nu-1} \rightarrow x_\nu (1 \leqq \nu \leqq n)$ is a singular edge of Γ, and as $a \rightarrow b$ is a regular edge of Γ we have

$$(x_{\nu-1}, x_\nu) \neq (a, b) \qquad (1 \leqq \nu \leqq n).$$

Hence, by definition of Γ', the relations (41) hold in Γ' as well as in Γ, and consequently $x \rightarrow y$ is a singular edge of Γ'.

If, on the other hand, $x \rightarrow y$ is a singular edge of Γ' then some cycle of the form (41) exists in Γ' and hence, a fortiori, in Γ. In other words, Γ and Γ' have the same system of singular edges. Therefore, remembering that $a \rightarrow b$ is true in Γ but not true in Γ', we find $\rho(\Gamma') = \rho(\Gamma) - 1$.

Now, according to our induction hypothesis, there is a T'-transform $\bar{f}$ of f''' which satisfies (5). $\bar{f}$ is a T'-transform with respect to Γ' and therefore, a fortiori, with respect to Γ. This completes the proof of Theorem 1 and, as we have seen, also of Theorem 2.

6. Proof of theorems 3 and 4

6.1 In this section we suppose that (2) and (4) hold.

Lemma 8. *Let $x' \equiv x''$. Then there are T''-transforms $f^{(1)}, f^{(2)}$, of f which satisfy $|f^{(1)}(x'')| \leqq |f(x')| \leqq |f^{(2)}(x'')|$,*

$$f^{(1)}(x) = f^{(2)}(x) = f(x) \quad (x \not\equiv x'),$$

$$|f^{(\lambda)}(x)| = |f(\phi^{(\lambda)}(x))| \qquad (1 \leqq \lambda \leqq 2; \qquad x \equiv x')$$

where, for each λ, $\phi^{(\lambda)}(x)$ effects a permutation of the points x of $\mathfrak{K}(x')$.

Proof. 1) Since $x'' \rightarrow\rightarrow x'$ there is a chain

$$x'' = x_o \rightarrow x_1 \rightarrow \cdots \rightarrow x_n = x'$$

where $n \geqq 0$. Then

$$x_\nu \equiv x' \quad (0 \leqq \nu \leqq n).$$

Let m be the smallest index which satisfies

$$0 \leqq m \leqq n; \qquad |f(x_m)| \leqq |f(x')|.$$

If $m = 0$ then $|f(x'')| \leqq |f(x')|$, and hence we may put $f^{(1)}(x) = f(x)$. Now suppose that $m > 0$. Then

$$|f(x_\nu)| > |f(x')| \geqq |f(x_m)| \qquad (0 \leqq \nu < m).$$

By (C) there is a set $A_{m-1} \subset f(x_{m-1})$ such that

$$|f(x_{m-1}) - A_{m-1}| = |f(x_m)|,$$
$$|f(x_m) + A_{m-1}| = |(f(x_{m-1}) - A_{m-1}) + A_{m-1}|.$$

Therefore

$$T^{x_{m-1}, x_m, A_{m-1}} f = f_1$$

is a T''-transformation, and

$$|f_1(x_{m-1})| = |f(x_m)| \leqq |f(x')|.$$

Now treat x_{m-1}, f_1 in the same way as x_m, f, etc., until, after m such T''-transformations, f is transformed into a function f_m which satisfies

$$|f_m(x_\gamma)| = \begin{cases} |f(x_m)| & (\nu = 0) \\ |f(x_{\nu-1})| & (0 < \nu \leqq m), \end{cases}$$
$$f_m(x) = f(x) \qquad (x \neq x_o, x_1, \cdots, x_m).$$

Hence we may put

$$f^{(1)}(x) = f_m(x).$$

2) Since $x' \to \to x''$ there is a chain

$$x' = y_o \to y_1 \to \cdots \to y_k = x''$$

where $k \geqq 0$. Let l be the largest index satisfying

$$0 \leqq l \leqq k; \qquad |f(y_l)| \geqq |f(x')|.$$

If $l = k$ then

$$|f(x'')| \geqq |f(x')|,$$

and we may put $f^{(2)}(x) = f(x)$. Now suppose that $l < k$. Then

$$|f(y_\nu)| < |f(x')| \leqq |f(y_l)| \qquad (l < \nu \leqq k).$$

Hence there is a set $B_l \subset f(y_l)$ such that

$$|f(y_l) - B_l| = |f(y_{l+1})|,$$
$$|f(y_{l+1}) + B_l| = |f(y_l)|,$$

and

$$T^{y_l, y_{l+1}, B_l} f = f_1'$$

is a T''-transformation,

$$|f_1'(y_{l+1})| = |f(y_l)| \geqq |f(x')|.$$

Now treat y_{l+1}, f_1' in the same way as y_l, f, etc., until, after $k - l$ such T''-transformations, f is transformed into a function f_{k-l}' satisfying

$$|f_{k-l}'(y_\nu)| = \begin{cases} |f(y_{\nu+1})| & (l \leqq \nu < k) \\ |f(y_l)| & (\nu = k), \end{cases}$$

$$f_{k-l}'(x) = f(x) \qquad (x \neq y_l, \quad y_{l+1}, \cdots, y_k).$$

Thus we may put

$$f^{(2)}(x) = f_{k-l}'(x).$$

6.2 LEMMA 9. *Suppose that $g(x)$ is decreasing, $\mathfrak{K}_o = \mathfrak{K}(x_o)$,*

$$|(f, \mathfrak{K}_o)| \geqq |(g, \mathfrak{K}_o)|. \tag{42}$$

Then there exists a T''-transform $\bar{f}$ of f satisfying

$$|\bar{f}(x)| \geqq |g(x)| \quad (x < \mathfrak{K}_o)$$

$$\bar{f}(x) = f(x) \qquad (x \not< \mathfrak{K}_o).$$

PROOF. If $x < \mathfrak{K}_o$ then $x \to\to x_o \to\to x$ and hence

$$|g(x)| \geqq |g(x_o)| \geqq |g(x)|,$$

$$|g(x)| = |g(x_o)| \qquad (x < \mathfrak{K}_o). \tag{43}$$

Define an operator O_4 as follows. O_4 is applicable to every f which satisfies (42) and

$$|f(x')| < |g(x')| \tag{44}$$

for at least one $x' < \mathfrak{K}_o$. We may assume that O_4 is applicable to the given function f. Choose one such x'. Then there exists $x'' < \mathfrak{K}_o - \{x'\}$ such that

$$|f(x'')| > |g(x_o)|. \tag{45}$$

For if

$$|f(x)| \leqq |g(x_o)| \qquad (x < \mathfrak{K}_o - \{x'\}) \tag{46}$$

were true then, by Lemma 2 and (43), (44), (46) we obtain a contradiction against (42). Choose a point $x'' < \mathfrak{K}_o - \{x'\}$ which satisfies (45). Then there is a chain

$$x'' = a_o \to a_1 \to \cdots \to a_n = x',$$

where $n > 0$ and all a_ν belong to $\mathfrak{K}_o$. Let ν_o be the largest index for which

$$0 \leqq \nu_o < n; \qquad |f(a_{\nu_o})| > |g(x_o)|.$$

Let ν_1 be the smallest index for which

$$\nu_o < \nu_1 \leqq n; \qquad |f(a_{\nu_1})| \neq |g(x_o)|.$$

Then

$$|f(a_\nu)| \begin{cases} = |g(x_o)| & (\nu_o < \nu < \nu_1) \\ < |g(x_o)| & (\nu = \nu_1). \end{cases}$$

In case we should have $\nu_o + 1 < \nu_1$ choose a set $A_o \subset f(a_{\nu_o})$ satisfying

$$|f(a_{\nu_o}) - A_o| = |g(x_o)|.$$

Then, by (A),

$$|f(a_{\nu_o+1}) + A_o| = |f(a_{\nu_o})|.$$

Hence

$$T^{a_{\nu_o}, a_{\nu_o+1}, A_o} f = f_1$$

is a T''-transformation. We have

$$|f_1(a_\nu)| = \begin{cases} |g(x_o)| & (\nu = \nu_o) \\ |f(a_{\nu_o})| & (\nu = \nu_o + 1) \end{cases}$$

$$f_1(x) = f(x) \qquad (x \neq a_{\nu_o}, a_{\nu_o+1}).$$

In case $\nu_o + 2 < \nu_1$, apply to f_1, $\nu_o + 1$ the same process as to f, ν_o, etc., until, after $r = \nu_1 - \nu_o + 1$ such transformations, we obtain a T''-transform f_r of f which satisfies

$$|f_r(a_\nu)| = \begin{cases} |f(a_{\nu_o})| & (\nu = \nu_1 - 1) \\ |g(x_o)| & (\nu_o \leqq \nu < \nu_1 - 1), \end{cases}$$

$$f_r(x) = f(x) \qquad (x \neq a_{\nu_o}, a_{\nu_o+1}, \cdots, a_{\nu_1-1}).$$

We put $f_o = f$, and then the result holds also for $r = 0$. Let

$$a_{\nu_1-1} = a; \qquad a_{\nu_1} = b; \qquad f_r = f'.$$

Then

$$a, b \prec \mathfrak{R}_o; \qquad a \rightarrow b;$$

$$|f'(a)| > |g(x_o)|; \qquad |f'(b)| < |g(x_o)|.$$

By (C) there is a set $A \subset f'(a)$ such that

$$|f'(a) - A| = |g(x_o)|.$$

Case 1. Suppose that $|f'(b) + A| \leqq |f'(a)|$. Then

$$T^{a,b,A} f' = f''$$

is a T''-transformation, and $|f''(a)| = |g(x_o)|$. In this case put, by definition, $O_4 f = f''$.

CASE 2. Suppose that $|f'(b) + A| > |f'(a)|$. Then

$$|f'(b)| < |g(x_o)| < |f'(a)| < |f'(b) + A|.$$

Hence an application of Lemma 3, with

$$m = 1; \qquad A_1 = f'(b); \qquad B_1 = g(x_o)$$

shows the existence of a set $A' \subset A$ satisfying

$$|f'(b) + A'| = |g(x_o)|.$$

Then

$$|f'(b) + A'| < |f'(a)|,$$

and therefore

$$T^{a,b,A'}f' = f''$$

is a T''-transformation. We have

$$|f''(b)| = |g(x_o)|.$$

Again define $O_4 f$ by putting

$$O_4 f = f''.$$

In either case the number of points x for which

$$|f''(x)| = |g(x_o)|$$

exceeds the number of points x for which

$$|f(x)| = |g(x_o)|.$$

Therefore O_4 can only be applied to f a bounded number of times, and after the last possible application we obtain a function $\bar{f}$ of the required type.

6.3 We now prove Theorem 3. Suppose that g is decreasing and that f satisfies (6), in addition to (2) and (4). We have to show the existence of a T''-transform $\bar{f}$ of f for which (5) is true.

CASE 1. Suppose that

$$\left\{\begin{array}{l} |f(a)| \leqq |f(b)| \\ \textit{whenever} \quad a \rightarrow \rightarrow b \not\equiv a. \end{array}\right. \tag{47}$$

(47) is, for instance, satisfied if $\rho(\Gamma) = O$.

We define an operator O_5 as follows. O_5 is applicable to every f satisfying (6) and (47) which is, moreover, such that

$$|f(x_o)| < |g(x_o)| \tag{48}$$

for at least one point x_o. We may assume that O_5 is applicable to the given function f. Choose one such x_o. Then

$$\mathfrak{K}_o = \mathfrak{K}(x_o)$$

is a section. For otherwise there would exist points x_1, x_2 such that

$$x_1 \rightarrow x_2 \equiv x_o \not\equiv x_1 .$$

Then

$$\mathfrak{A}_o = [\{x_o\}] - \mathfrak{K}_o$$

is a non-empty section. For, first of all, $x_1 \prec \mathfrak{K}_o$. Also, if

$$x \rightarrow y \prec \mathfrak{A}_o$$

then

$$x \rightarrow y \rightarrow \rightarrow x_o \not\equiv y, \qquad x \rightarrow \rightarrow x_o . \tag{49}$$

But $x \equiv x_o$ is impossible. For this would imply

$$x \rightarrow y \rightarrow \rightarrow x_o \rightarrow \rightarrow x$$

which would already contradict (49). Hence

$$x \rightarrow \rightarrow x_o \not\equiv x, \tag{50}$$

i.e., $x \prec \mathfrak{A}_o$. Therefore $\mathfrak{A}_o$ is a section.

Every point x of $\mathfrak{A}_o$ satisfies (50). Therefore, by (47), (48) and the fact that g is decreasing,

$$| f(x) | \leqq | f(x_o) | < | g(x_o) | \leqq | g(x) | \qquad (x \prec \mathfrak{A}_o).$$

Now Lemma 2 shows that

$$| (f, \mathfrak{A}_o) | < | (g, \mathfrak{A}_o) |$$

which contradicts (6). This contradiction was deduced from the assumption that $\mathfrak{K}_o$ was no section. Therefore $\mathfrak{K}_o$ is, in fact, a section and, in consequence, satisfies (42).

By Lemma 9 we can find a T''-transform f' of f which is such that

$$| f'(x) | \geqq | g(x) | \qquad (x \prec \mathfrak{K}_o)$$
$$f'(x) = f(x) \qquad (x \not\prec \mathfrak{K}_o).$$

Define O_5 by putting

$$O_5 f = f'.$$

There are more points x for which

$$| f'(x) | \geqq | g(x) |$$

than points x for which

$$| f(x) | \geqq | g(x) |.$$

Hence O_6 can only be applied to f a bounded number of times, and after the last possible application we obtain a function $\bar{f}$ with the required properties.

CASE 2. Suppose that there are points a, b for which

$$a \to \to b \not\equiv a, \tag{51}$$

$$|f(a)| > |f(b)|. \tag{52}$$

In this case the proof is based on the following principle. By means of a certain process the problem presented by f, Γ will be reduced to an analogous problem f^*, Γ^*, and we shall have either

$$\Gamma^* = \Gamma; \qquad s(f^*, \Gamma^*) > s(f, \Gamma) \tag{i}$$

or

$$\rho(\Gamma^*) < \rho(\Gamma). \tag{ii}$$

If such reduction process has been established then the given problem can be reduced to one falling under case one, i.e. where (47) holds. For if this were not so the reduction process could be applied infinitely often and would lead to a sequence of pairs $f^{(n)}$, $\Gamma^{(n)}$ which is such that, for every $n = 1, 2, \cdots$, either

$$\Gamma^{(n+1)} = \Gamma^{(n)}; \qquad s(f^{(n+1)}, \Gamma^{(n+1)} > s(f^{(n)}, \Gamma^{(n)}) \tag{i}$$

or

$$\rho(\Gamma^{(n+1)}) < \rho(\Gamma^{(n)}). \tag{ii}$$

In particular, $\rho(\Gamma^{(n+1)} \leqq \rho(\Gamma^{(n)})$ for all n. Since $s(f, \Gamma)$ does not exceed a constant $C(\Gamma)$ independent of f—e.g. $C(\Gamma) = N^3$ where N is the number of points of Γ—it follows that only a finite number of consecutive integers n can satisfy (i). Therefore $\rho(\Gamma^{(n)}) \to -\infty$ as $n \to \infty$, which is absurd.

In fact, in case (ii) $\Gamma^{(n+1)}$ is obtained from $\Gamma^{(n)}$ by removing one edge, and hence it follows that the number of possible applications of the reduction process is less than a constant dependent only on Γ.

In order to define the reduction process we start by noting that, according to (51), there exists a chain

$$a = a_0 \to a_1 \to \cdots \to a_m = b.$$

Then $m > 0$ since otherwise $a = b$ which contradicts (51). Let

$$\mu_1, \mu_2, \cdots, \mu_k$$

be those indices μ for which

$$1 \leqq \mu \leqq m; \qquad a_{\mu-1} \not\equiv a_\mu,$$

and let

$$1 \leqq \mu_1 < \mu_2 < \cdots < \mu_k \leqq m.$$

Then $k > 0$ since otherwise (51) would again be violated. Put

$$a_{\mu_\kappa - 1} = x_\kappa ; \qquad a_{\mu_\kappa} = y_\kappa \qquad\qquad (1 \leqq \kappa \leqq k),$$

$$y_o = a; \qquad x_{k+1} = b.$$

Then

$$y_o \equiv x_1 \rightarrow y_1 \equiv x_2 \rightarrow y_2 \equiv \cdots \equiv x_k \rightarrow y_k \equiv x_{k+1} ,$$

$$x_\kappa \not\equiv y_\kappa \qquad\qquad (1 \leqq \kappa \leqq k).$$

Not all inequalities

$$| f(y_o) | \leqq | f(y_1) | \leqq | f(y_2) | \leqq \cdots \leqq | f(y_{k-1}) | \leqq | f(x_{k+1}) |$$

are true. For this would contradict (52). Hence either there is a number κ_o satisfying

(53) $$1 \leqq \kappa_o < k; \qquad | f(y_{\kappa_o - 1}) | > | f(y_{\kappa_o}) |$$

or we have

(54) $$| f(y_{k-1}) | > | f(b) | .$$

In case (53) we put

$$a' = y_{\kappa_o - 1} ; \qquad a'' = x_{\kappa_o} ; \qquad b' = b'' = y_{\kappa_o} .$$

In case (54) we put

$$a' = y_{k-1} ; \qquad a'' = x_k ; \qquad b' = x_{k+1} ; \qquad b'' = y_k .$$

Then in either case

$$a' \equiv a'' \rightarrow b'' \equiv b'; \qquad a'' \not\equiv b'',$$

$$| f(a') | > | f(b') | .$$

Apply to $f(x)$ Lemma 8, with $x' = a'$; $x'' = a''$. We obtain a T''-transform f_1 of f which satisfies

$$| f_1(a'') | \geqq | f(a') | ,$$

(55) $$| f_1(x) | = | f(\phi^{(2)}(x)) | \quad \textit{for} \quad x \equiv a'$$

(56) $$f_1(x) = f(x) \qquad \textit{for} \quad x \not\equiv a'$$

where $\phi^{(2)}(x)$ is a permutation of the points of $\mathfrak{K}(a')$. Now apply Lemma 8 to $f_1(x)$, with $x' = b'$; $x'' = b''$. We obtain a T''-transform f_2 of f_1 which satisfies

$$| f_2(b'') | \leqq | f_1(b') | ,$$

$$| f_2(x) | = | f_1(\phi^{(1)}(x)) | \quad \textit{for} \quad x \equiv b',$$

$$f_2(x) = f_1(x) \qquad \textit{for} \quad x \not\equiv b'$$

where $\phi^{(1)}(x)$ is a permutation of the points of $\mathfrak{K}(b')$.

Let $\mathfrak{A}$ be any section. If $a' < \mathfrak{A}$ then $\mathfrak{K}(a') \subset \mathfrak{A}$ and hence, by (56),

$$(f_1, \mathfrak{A}) = (f_1, \mathfrak{G}) - (f_1, \mathfrak{G} - \mathfrak{A}) = (f, \mathfrak{G}) - (f, \mathfrak{G} - \mathfrak{A}) = (f, \mathfrak{A}).$$

On the other hand, if $a' \nless \mathfrak{A}$ then $\mathfrak{A}\mathfrak{K}(a') = 0$ and, by (56),

$$(f_1, \mathfrak{A}) = (f, \mathfrak{A}). \tag{57}$$

Thus (57) holds for every $\mathfrak{A}$. In the same way it follows that

$$(f_2, \mathfrak{A}) = (f_1, \mathfrak{A}) \quad \textit{for all } \mathfrak{A}.$$

Therefore

$$|(f_2, \mathfrak{A})| = |(f, \mathfrak{A})| \geqq |(g, \mathfrak{A})| \tag{58}$$

for all $\mathfrak{A}$. Furthermore,

$$|f_2(x)| = |f(\phi(x))| \quad \textit{for all } x$$

where

$$\phi(x) = \begin{cases} \phi^{(2)}(x) & (x \equiv a') \\ \phi^{(1)}(x) & (x \equiv b') \\ x & (x \not\equiv a', b'). \end{cases}$$

On making use of the various properties of our points and functions we see that

$$|f_2(a'')| = |f_1(a'')| \geqq |f(a')| > |f(b')| = |f_1(b')| \geqq |f_2(b'')|. \tag{59}$$

Therefore, by (C), we can choose a set $A \subset f_2(a'')$ such that

$$|f_2(a'') - A| = |f_2(b'')|.$$

Then, by (A),

$$|f_2(b'') + A| = |f_2(a'')|.$$

Hence

$$T^{a'',b'',A} f_2 = f_3$$

is a T''-transformation.

Case (i). Assume that

$$|(f_3, \mathfrak{A})| \geqq |(g, \mathfrak{A})| \qquad \textit{for all } \mathfrak{A}.$$

We have

$$|f_3(a'')| = |f_2(b'')|; \qquad |f_3(b'')| = |f_2(a'')|,$$
$$f_3(x) = f_2(x) \qquad (x \neq a'', b'').$$

For the sake of simplicity we write $p(\bar{x})$, $q(f, \bar{x})$, $s(f)$ instead of $p(\bar{x}, \Gamma)$, $q(f, \bar{x}, \Gamma)$, $s(f, \Gamma)$ respectively.

From (56), (55) and the definition of $q(f, \bar{x})$ it is obvious that

$$q(f_1, \bar{x}) = q(f, \bar{x}) \qquad (\bar{x} \not\equiv a').$$

Also,

$$p(x) = p(y) \qquad \textit{for } x \equiv y,$$
$$q(f_1, \bar{x}) = q(f, \phi^{(2)}(\bar{x})) \qquad \textit{for all } \bar{x} \equiv a'.$$

Therefore

$$s(f_1) = \sum_{\bar{x} \neq a'} p(\bar{x})q(f_1, \bar{x}) + \sum_{\bar{x} = a'} p(\bar{x})q(f_1, \bar{x})$$

$$= \sum_{\bar{x} \neq a'} p(\bar{x})q(f, \bar{x}) + \sum_{\bar{x} = a'} p(\phi^{(2)}(\bar{x}))q(f, \phi^{(2)}(\bar{x})) = s(f).$$

Similarly, using b', $\phi^{(1)}$ instead of a', $\phi^{(2)}$, we see that

$$s(f_2) = s(f_1).$$

Thus

(60) $$s(f_2) = s(f).$$

Furthermore,

$$q(f_3, \bar{x}) = q(f_2, \bar{x}) \qquad \textit{for } \bar{x} \neq a'', b'',$$

$$p(a'') < p(b'')$$

and, by (59),

$$q(f_3, a'') = q(f_2, b'') < q(f_2, a'') = q(f_3, b'').$$

Hence

$$s(f_3) - s(f_2)$$

$$= p(a'')q(f_3, a'') + p(b'')q(f_3, b'') - p(a'')q(f_2, a'') - p(b'')q(f_2, b'')$$

$$= (p(b'') - p(a''))(q(f_2, a'') - q(f_2, b'')) > 0,$$

and finally, by (60),

$$s(f_3, \Gamma) > s(f, \Gamma).$$

This is case (i) of the reduction process mentioned on p. 34, with

$$f^* = f_3.$$

CASE (ii). Suppose there exists a section $\mathfrak{A}_1$ for which

$$|(f_3, \mathfrak{A}_1)| < |(g, \mathfrak{A}_1)|.$$

Then

$$a'' \prec \mathfrak{A}_1; \qquad b'' \not\prec \mathfrak{A}_1.$$

As in the proof of Theorem 1, let

$$\mathfrak{A}_1, \mathfrak{A}_2, \cdots, \mathfrak{A}_m$$

be all sections for which

$$a'' \prec \mathfrak{A}_\mu; \qquad b'' \not\prec \mathfrak{A}_\mu.$$

Put

$$(f_2, \mathfrak{A}_\mu) = A_\mu; \qquad (g, \mathfrak{A}_\mu) = B_\mu \qquad (1 \leqq \mu \leqq m).$$

Then

$$|A_1 - A| = |(f_2, \mathfrak{A}_1) - A| = |(f_3, \mathfrak{A}_1)| < |(g, \mathfrak{A}_1)| = B_1,$$

$$|(A_\mu - A) + A| = |A_\mu| \geqq |B_\mu| \qquad (1 \leqq \mu \leqq m) \quad (\text{by } (58)).$$

Therefore, by Lemma 3, there is a set $A' \subset A$ and an index μ_o $(1 \leqq \mu_o \leqq m)$ such that

$$| (A_\mu - A) + A' | \begin{cases} \geqq | B_\mu | & (1 \leqq \mu \leqq m) \\ = | B_\mu | & (\mu = \mu_o). \end{cases}$$

Then

$$T^{a'', b'', A - A'} f_2 = f_4$$

is a T''-transformation, and we have

$$| (f_4 , \mathfrak{A}_\mu) | \begin{cases} \geqq | (g, \mathfrak{A}_\mu) | & (1 \leqq \mu \leqq m) \\ = | (g, \mathfrak{A}_\mu) | & (\mu = \mu_o). \end{cases}$$

If $\mathfrak{A}$ is a section which does not occur among the $\mathfrak{A}_\mu$ then

$$\textit{either} \quad a'', b'' \prec \mathfrak{A} \quad \textit{or} \quad a'', b'' \not\prec \mathfrak{A},$$

and hence in any case

$$(f_4 , \mathfrak{A}) = (f_2 , \mathfrak{A}).$$

Therefore

$$| (f_4 , \mathfrak{A}) | \geqq | (g, \mathfrak{A}) | \qquad \textit{for all } \mathfrak{A},$$

$$| (f_4 , \mathfrak{A}_o) | = | (g, \mathfrak{A}_o) |$$

when

$$\mathfrak{A}_o = \mathfrak{A}_{\mu_o} .$$

Now apply Lemma 5, with

$$a_o = a''; \qquad b_o = b''.$$

Then, in the notations of Lemma 5,

$$| (f_4 , \mathfrak{A}') | \geqq | (g, \mathfrak{A}') |$$

for every section $\mathfrak{A}'$ of Γ', and we have

$$\rho(\Gamma') = \rho(\Gamma) - 1.$$

This is case (ii) of the reduction process, with

$$\Gamma^* = \Gamma'; \qquad f^* = f_4 .$$

Theorems 3 and 4 are proved.

7. Proof of theorems 5 and 6

7.1 We now consider the case of two arbitrary functions $f(x)$, $g(x)$, not necessarily satisfying (4). But (2) is required. It will be shown that the transformation problem of Theorem 5 is identical with that of Theorem 1; if the latter is applied to the functions $F(\mathfrak{B})$, $G(\mathfrak{B})$ (defined on p. 10) and a certain graph Γ^*. The condition (14) is, in fact, identical with (6), applied to F, G, Γ^*.

7.2 It is obvious from the definition of $F(\mathfrak{B})$ that

$$(61) \qquad F(\mathfrak{B}_1)F(\mathfrak{B}_2) = 0 \qquad \textit{if } \mathfrak{B}_1 \neq \mathfrak{B}_2 .$$

To prove this formally we may assume that there is a point x_o such that $x_o \prec \mathfrak{B}_1$; $x_o \nprec \mathfrak{B}_2$. Then by (10)

$$F(\mathfrak{B}_1)F(\mathfrak{B}_2) \subset f(x_o)(S - f(x_o)) = 0.$$

Denote by Γ^* a graph whose "points" are all sets $\mathfrak{B} \subset \mathfrak{G}$. The edges of Γ^* will be defined later. The symbol $\rightarrow$ refers to Γ. Edges of Γ^* will be denoted by the symbol $\xrightarrow{*}$. More generally, the addition of a star * to any of our symbols indicates that this symbol refers to Γ^*.

Suppose that

$$a \rightarrow b; \qquad A \subset f(a),$$
$$T^{a,b,A}f = f_1 .$$

We want to investigate the relation between the functions $F(\mathfrak{B})$ and $F_1(\mathfrak{B})$.

We have, by the distributive law of multiplication,

$$A = Af(a) \prod_{x \prec \mathfrak{G} - \{a\}} (f(x) + (S - f(x))) = \sum_{a \prec \mathfrak{C}} AF(\mathfrak{C}).$$

Hence, in view of (61), the transformation $T^{a,b,A}$ is equivalent to the succession of transformations

$$T^{a,b,AF(\mathfrak{C})}$$

where $\mathfrak{C}$ ranges over all sets which contain the point a. It is therefore sufficient for our purpose to assume that

$$0 \neq A \subset F(\mathfrak{C})$$

where $\mathfrak{C}$ is some fixed set and

$$a \prec \mathfrak{C}.$$

Consider any set $\mathfrak{B}$ satisfying

$$a, b \nprec \mathfrak{B}.$$

Put

$$B = \prod_{x \prec \mathfrak{B}} f(x) \prod_{y \prec \mathfrak{B} + \{a,b\}} (S - f(x)). \tag{62}$$

Then

$$F_1(\mathfrak{B}) = B(S - (f(a) - A))(S - (f(b) + A))$$
$$= B((S - f(a)) + A)(S - f(b))(S - A) = F(\mathfrak{B})(S - A) = F(\mathfrak{B}). \tag{63}$$

The last equation follows from

$$AF(\mathfrak{B}) \subset F(\mathfrak{C})F(\mathfrak{B})$$

and (61). Similarly,

$$F_1(\mathfrak{B} + \{a\}) = B(f(a) - A)(S - (f(b) + A))$$
$$= Bf(a)(S - A)(S - f(b)) = F(\mathfrak{B} + \{a\})(S - A)$$

$$(64) \qquad = \begin{cases} F(\mathfrak{B} + \{a\}) - A & \textit{when } \mathfrak{B} + \{a\} = \mathfrak{C} \\ F(\mathfrak{B} + \{a\}) & \textit{otherwise.} \end{cases}$$

$$F_1(\mathfrak{B} + \{b\}) = B(f(b) + A)(S - (f(a) - A)) = Bf(b)(S - f(a)) + AB$$

$$(65) \qquad = F(\mathfrak{B} + \{b\}) + AB = \begin{cases} F(\mathfrak{B} + \{b\}) + A & \textit{when } A \subset B \\ F(\mathfrak{B} + \{b\}) & \textit{otherwise.} \end{cases}$$

$$F_1(\mathfrak{B} + \{a, b\}) = B(f(a) - A)(f(b) + A) = Bf(a)(S - A)f(b)$$

$$(66) \quad = F(\mathfrak{B} + \{a, b\})(S - A) = \begin{cases} F(\mathfrak{B} + \{a, b\}) - A & \textit{when } \mathfrak{B} + \{a, b\} = \mathfrak{C} \\ F(\mathfrak{B} + \{a, b\}) & \textit{otherwise.} \end{cases}$$

I assert that $A \subset B$ if, and only if,

$$(67) \qquad \mathfrak{B} + \{b\} = (\mathfrak{C} - \{a\}) + \{b\}.$$

First we notice that $A \subset B$ if, and only if, $F(\mathfrak{C}) \subset B$. Now it follows from (62) and the distributive law of multiplication that

$$B = B(f(a) + (S - f(a)))(f(b) + (S - f(b)))$$
$$= F(\mathfrak{B}) + F(\mathfrak{B} + \{a\}) + F(\mathfrak{B} + \{b\}) + F(\mathfrak{B} + \{a, b\}).$$

Hence $A \subset B$ is equivalent to

$$(68) \qquad \mathfrak{C} = \mathfrak{B} + \mathfrak{C}' \quad \text{where} \quad \mathfrak{C}' = \{a\} \text{ or } \{a, b\}$$

Since $a \prec \mathfrak{C}$, (68) is the same as

$$\mathfrak{B} \subset \mathfrak{C} - \{a\} \subset \mathfrak{B} + \{b\}$$

which, in turn, is equivalent to (67).

If we examine the equations (63)–(66), bearing in mind that $A \subset B$ is equivalent to (67), we find that they can be summarised as follows.

$$F_1(\mathfrak{B}_1) = \begin{cases} F(\mathfrak{B}_1) - A & \textit{if } \mathfrak{B}_1 = \mathfrak{C} \\ F(\mathfrak{B}_1) + A & \textit{if } \mathfrak{B}_1 = (\mathfrak{C} - \{a\}) + \{b\} \\ F(\mathfrak{B}_1) & \textit{otherwise.} \end{cases}$$

Hence the appropriate definition of edges in Γ^*—the graph whose "points" are all sets $\mathfrak{B}$ of points of Γ—is

$$(69) \qquad \mathfrak{C} \xrightarrow{*} (\mathfrak{C} - \{a\}) + \{b\}$$

where a, b, $\mathfrak{C}$ satisfy

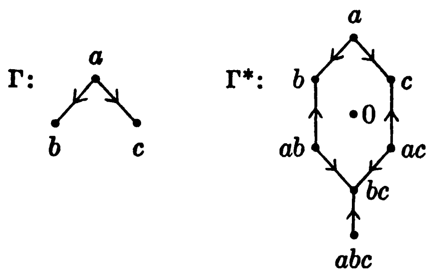
$$a \rightarrow b; \qquad a \prec \mathfrak{C}$$

and are otherwise arbitrary. For if these are the edges of Γ^*, and we actually adopt this definition, then every T-transformation applied to $f(x)$ and the graph Γ has an effect upon the function $F(\mathfrak{B})$ which is the same as that of a succession

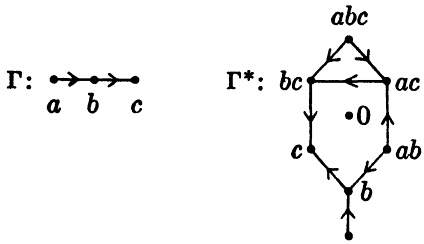
of T-transformations applied to $F(\mathfrak{B})$ and the graph Γ^*. Vice versa, our analysis shows that every T-transformation applied to $F(\mathfrak{B})$ and Γ^* is equivalent, in its effect upon $f(x)$, to a T-transformation applied to $f(x)$ and Γ.

The following diagrams serve to illustrate the connection between Γ and Γ^*. Points of Γ^* which are denoted by ab, abc etc. should really be called $\{a, b\}$, $\{a, b, c\}$, etc. The point of Γ^* called 0 corresponds to the set $\mathfrak{B} = 0$. It is always an isolated point of Γ^*.

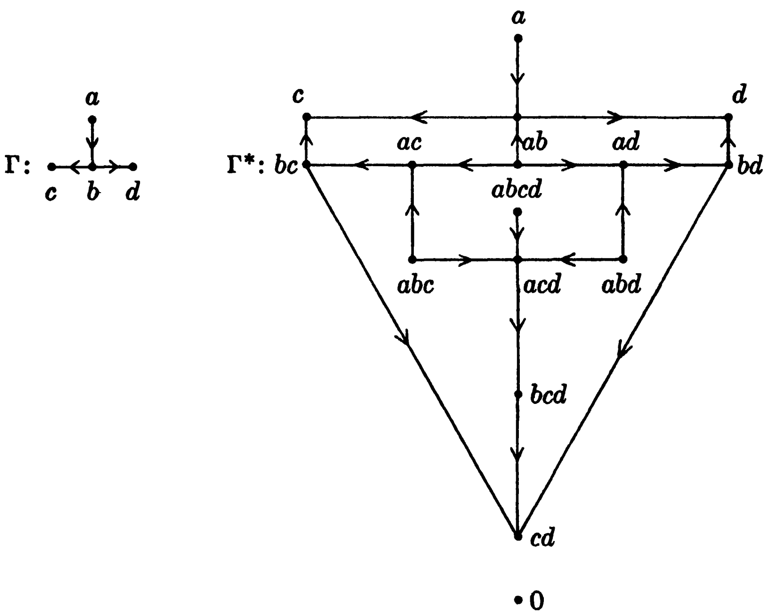
1)



2)



3)



Even very simple graphs Γ give rise to quite complicated and interesting graphs Γ^*. In example 3) a few edges, e.g. $abcd \xrightarrow{*} bcd$ have been omitted in Γ^*, but the validity of relations $\mathfrak{B} \overset{*}{\rightarrow\!\!\!\rightarrow} \mathfrak{C}$ is not affected by these changes. Therefore the sections of the reproduced graph Γ^* are the same as those of the true graph Γ^*. Hence the meaning of condition (6) and thus the existence or non-existence of any T-transform mentioned in our theorems is not affected.

7.3 In order to complete the proof of theorems 5 and 6, all we have to do is to find the sections of Γ^* and to verify that (14) is the same as (6) if the latter is applied to Γ^* instead of Γ. We have to note here that

$$(f, \mathfrak{S}) = \sum_{\mathfrak{B}} F(\mathfrak{B}) = (F, \mathfrak{S}^*)^*,$$

and that therefore (8) serves for both the problem on Γ and that on Γ^*.

The definition (69) can be expressed as follows. Let x_o be an arbitrary point of Γ, and $\phi_o(x_o)$ a point such that

$$x_o \rightarrow \phi_o(x_o).$$

Put

$$\phi_o(x) = x \qquad \textit{for } x \neq x_o .$$

Then, if $x_o < \mathfrak{B}_o$, we have

$$\mathfrak{B}_o \xrightarrow{*} \sum_{x < \mathfrak{B}_0} \{\phi_o(x)\}.$$

Therefore, for any $\mathfrak{B}$,

$$\mathfrak{B} \overset{*}{\rightarrow\!\!\!\rightarrow} \sum_{x < \mathfrak{B}} \{\phi_o(x)\}.$$

More generally, let $\phi(x)$ be defined for all x and denote a point of Γ for which

$$x \rightarrow\!\!\!\rightarrow \phi(x) \qquad \textit{for all } x. \tag{70}$$

Then every $\mathfrak{B}$ satisfies

$$\mathfrak{B} \overset{*}{\rightarrow\!\!\!\rightarrow} \sum_{x < \mathfrak{B}_0} \{\phi(x)\},$$

and in this way we obtain all relations $\mathfrak{B} \overset{*}{\rightarrow\!\!\!\rightarrow} \mathfrak{C}$.

7.4 We now want to determine all sections of Γ^* which can be written in the form $[\{\mathfrak{B}\}]^*$, i.e. which are "generated" by a single point $\mathfrak{B}$ of Γ^*. By (9) any arbitrary section of Γ^* is a sum of such special sections. We are given a set $\mathfrak{B}$, and we have to find all sets $\mathfrak{C}$ for which

$$\mathfrak{C} \overset{*}{\rightarrow\!\!\!\rightarrow} \mathfrak{B}. \tag{71}$$

In what follows the set $\mathfrak{B}$ is fixed. Let $\psi_{\mathfrak{B}}(x)$ be a function of the type defined on p. 11. Let $\mathfrak{C}$ be any set satisfying

$$\sum_{x < \mathfrak{B}} \{\psi_{\mathfrak{B}}(x)\} \subset \mathfrak{C} \subset [\mathfrak{B}]. \tag{72}$$

I am going to show now that sets $\mathfrak{C}$ which satisfy (72) for some $\psi_{\mathfrak{B}}(x)$, and no other sets, have the property (71).

1) Assume that (72) is true for a certain function $\psi_{\mathfrak{B}}(x)$ and a set $\mathfrak{C}$. Then we define a function $\phi(x)$ as follows.

$$\phi(x) = x \qquad \textit{for } x \not\prec \mathfrak{C}$$

$$\phi(\psi_{\mathfrak{B}}(y)) = y \qquad \textit{for } y \prec \mathfrak{B}.$$

Finally, if x belongs to the set

$$\mathfrak{C} - \sum_{y \prec \mathfrak{B}} \{\psi_{\mathfrak{B}}(y)\}$$

then, by (72),

$$x \prec [\mathfrak{B}].$$

Hence, by definition of $[\mathfrak{B}]$,

$$x \longrightarrow\!\!\longrightarrow x'$$

for at least one x' of $\mathfrak{B}$. Choose one such x' and put

$$\phi(x) = x'.$$

It follows from the properties of $\psi_{\mathfrak{B}}$ that the conditions imposed upon ϕ do not contradict each other and that ϕ satisfies (70). Furthermore,

$$\sum_{x \prec \mathfrak{C}} \{\phi(x)\} = \mathfrak{B}.$$

Therefore (71) follows.

2) Let us now suppose that a set $\mathfrak{C}$ satisfies (71). Then, by definition of $\xrightarrow{*}\!\!\rightarrow$, there exists a function $\phi(x)$ with property (70) for which

$$\sum_{x \prec \mathfrak{C}} \{\phi(x)\} = \mathfrak{B}. \tag{73}$$

Hence, given any point y of $\mathfrak{B}$, there is at least one $\bar{x}$ of $\mathfrak{C}$ satisfying $\phi(\bar{x}) = y$. Choose one such $\bar{x}$ and put

$$\psi(y) = \bar{x}.$$

Then

$$\psi(y) \prec \mathfrak{C}; \qquad \phi(\psi(y)) = y.$$

Clearly,

$$\psi(y_1) \neq \psi(y_2) \qquad \textit{if } y_1 \neq y_2$$

and, by (70),

$$\psi(y) \longrightarrow\!\!\longrightarrow y$$

when y, y_1, y_2 are in $\mathfrak{B}$. This means that $\psi(y)$ is one of those functions which may be denoted by $\psi_{\mathfrak{B}}(y)$. Also, in view of (70), (73), every x of $\mathfrak{C}$ belongs to $[\mathfrak{B}]$. Hence (72) holds.

7.5 Put

$$A = (F, [\{\mathfrak{B}\}]^*)^*.$$

From a remark made at the beginning of 7.4 it is obvious that our proof will be completed if we have established that, for every $\mathfrak{B} \neq 0$, A is equal to $F^*(\mathfrak{B})$ as defined by (12). For then (14) will be identical with (6), the latter applied to F, G, Γ^*.

We have, in fact,

$$A = \sum_{\mathfrak{C} \to \overset{*}{\to} \mathfrak{B}} F(\mathfrak{C}) = \sum_{\psi_{\mathfrak{B}}} \sum_{\mathfrak{C}_\psi} F(\mathfrak{C}_\psi), \tag{74}$$

where $\mathfrak{C}_\psi$, for every fixed $\psi_{\mathfrak{B}}$, ranges over all sets $\mathfrak{C}$ which satisfy (72). Put

$$\sum_{x < \mathfrak{B}} \{\psi_{\mathfrak{B}}(x)\} = \mathfrak{B}_\psi .$$

Then, for every fixed $\psi_{\mathfrak{B}}$,

$$\begin{aligned}
\sum_{\mathfrak{C}_\psi} F(\mathfrak{C}_\psi) &= \sum_{\mathfrak{C}_\psi} \Big(\prod_{x < \mathfrak{C}_\psi} f(x)\Big)\Big(\prod_{y \text{ not in } \mathfrak{C}_\psi} (S - f(y))\Big) \\
&= \sum_{\mathfrak{C}_\psi} \Big(\prod_{x < \mathfrak{B}_\psi} f(x)\Big)\Big(\prod_{x < \mathfrak{C}_\psi - \mathfrak{B}_\psi} f(x)\Big)\Big(\prod_{y < [\mathfrak{B}] - \mathfrak{C}_\psi} (S - f(y))\Big)\Big(\prod_{y \text{ not in } [\mathfrak{B}]} (S - f(y))\Big) \\
&= \Big(\prod_{x < \mathfrak{B}_\psi} f(x)\Big)\Big(\prod_{y \text{ not in } \mathfrak{B}} (S - f(y))\Big) \sum_{\mathfrak{B}_\psi \subset \mathfrak{C} \subset [\mathfrak{B}]} \Big(\prod_{x < \mathfrak{C} - \mathfrak{B}_\psi} f(x)\Big)\Big(\prod_{y < [\mathfrak{B}] - \mathfrak{C}} (S - f(y))\Big) \\
&= \Big(\prod_{x < \mathfrak{B}} f(\psi_{\mathfrak{B}}(x))\Big)\Big(\prod_{y \text{ not in } [\mathfrak{B}]} (S - f(x))\Big)S.
\end{aligned} \tag{75}$$

The sum in (75) has the value S in view of the distributive law. If we substitute in (74) we find that

$$A = \sum_{\psi_{\mathfrak{B}}} \Big(\prod_{x < \mathfrak{B}} f(\psi_{\mathfrak{B}}(x))\Big)\Big(\prod_{y \text{ not in } [\mathfrak{B}]} (S - f(y))\Big) = F^*(\mathfrak{B}).$$

8. Necessity and independence of the hypotheses of the theorems

8.1 Let Γ_o be the graph with $\mathfrak{G}_o = \{a, b, c\}$ whose edges are $a \to b$; $a \to c$. Assume that a measure $|A|$ is defined in S but that none of the properties (M), (A), (C), (D) are known to hold. Suppose that the following version of Theorem 2 holds.

Assumption α. *Given any functions* $f(x)$, $g(x)$, *for* $x < \mathfrak{G}_o$, *and supposing that*

$$f(x)f(y) = g(x)g(y) = 0 \qquad \textit{for } x \neq y$$

Then a necessary and sufficient condition for f *to have a* T-*transform* $\bar{f}$ *satisfying*

$$|\bar{f}(x)| = |g(x)| \qquad \textit{for all } x \tag{76}$$

is that

(77) $$|(f, \mathfrak{A})| \geqq |(g, \mathfrak{A})| \qquad \text{for all } \mathfrak{A} \text{ of } \Gamma_o$$

and

(78) $$|(f, \mathfrak{G}_o)| = |(g, \mathfrak{G}_o)|.$$

We are going to prove now that, under these circumstances, the measure has the properties (M), (A), (C), (D) and, moreover, that S contains only sets of finite measure. Thus all our assumptions concerning the measure are necessary for the truth of Theorem 2. The slight discrepancy due to the fact that here both, $f(x)$ and $g(x)$, are of finite measure while in Theorem 2 only $|g(x)| < \infty$, can be removed.

The non-empty sections of Γ_o are

$$\{a\}, \qquad \{a, b\}, \qquad \{a, c\}, \qquad \{a, b, c\}.$$

For simplicity a symbol

$$f = \begin{pmatrix} B, & & C \\ & A & \end{pmatrix}$$

stands for the definition of a function $f(x)$, namely for

$$f(a) = A; \qquad f(b) = B; \qquad f(c) = C.$$

1) Let $A \subset B$. Put

$$f = \begin{pmatrix} 0, & & 0 \\ & B & \end{pmatrix}; \qquad g = \begin{pmatrix} B - A, & & 0 \\ & A & \end{pmatrix}.$$

Then $T^{a,b,B-A} f = g$. Therefore, by α,

$$|B| = |f(a) + f(c)| \geqq |g(a) + g(c)| = |A|.$$

Hence property (M) holds.

2) Let

$$|A_1| = |B_1|; \qquad |A_2| = |B_2|; \qquad A_1A_2 = B_1B_2 = 0.$$

Put

$$f = \begin{pmatrix} A_1, & & A_2 \\ & 0 & \end{pmatrix}; \qquad g = \begin{pmatrix} B_1, & & B_2 \\ & 0 & \end{pmatrix}.$$

Then

$$T^{a,b,0} f = f; \qquad |f(x)| = |g(x)| \qquad \text{for all } x.$$

Hence, by α,

$$|A_1 + A_2| = |(f, \mathfrak{G}_o)| = |(g, \mathfrak{G}_o)| = |B_1 + B_2|.$$

Therefore (A) holds.

3) Let

$$|A| \geqq |B_1|; \qquad A - AB_1 = A_1; \qquad B_1 - AB_1 = B_2.$$

Put

$$f = \begin{pmatrix} B_2, & 0 \\ & A \end{pmatrix}; \qquad g = \begin{pmatrix} A_1, & 0 \\ & B_1 \end{pmatrix}.$$

Then (77) and (78) are satisfied. Hence, by α, there exist sets C, D satisfying

$$CD = 0; \qquad C + D \subset A$$

which are such that (76) holds when

$$\mathfrak{f} = \begin{pmatrix} B_2 + C, & D \\ A - (C + D) \end{pmatrix}.$$

Then

$$|A - (C + D)| = |\mathfrak{f}(a)| = |g(a)| = |B_1|.$$

This establishes (C).

4) Let

$$|A| \geqq |B_1 + B_2|; \qquad B_1B_2 = 0.$$

By (C) there is $A' \subset A$ such that $|A'| = |B_1 + B_2|$.
Put

$$f = \begin{pmatrix} 0, & 0 \\ & A' \end{pmatrix}; \qquad g = \begin{pmatrix} B_1, & B_2 \\ & 0 \end{pmatrix}.$$

Then, using (M), we deduce (77), (78). Hence, by α, there exist sets A_1, A_2 satisfying

$$A_1A_2 = 0; \qquad A_1 + A_2 \subset A'$$

which are such that (76) holds for

$$\mathfrak{f} = \begin{pmatrix} A_1, & A_2 \\ A' - (A_1 + A_2) \end{pmatrix}.$$

Then

$$|A_1| = |\mathfrak{f}(b)| = |g(b)| = |B_1|; \qquad |A_2| = |B_2|.$$

Therefore (D) holds.

5) Let

$$A' \subset A; \qquad |A'| = |A|.$$

Put

$$f = \begin{pmatrix} A - A', & 0 \\ & A' \end{pmatrix}; \qquad g = \begin{pmatrix} 0, & 0 \\ & A \end{pmatrix}.$$

Then (77) and (78) hold. There are sets B, C with

$$BC = 0; \qquad B + C \subset A',$$

which are such that (76) is true for

$$\bar{f} = \begin{pmatrix} (A - A') + B, & C \\ & A' - (B + C) \end{pmatrix}.$$

Therefore, by (M),

$$|A - A'| \leqq |(A - A') + B| = |\bar{f}(b)| = |g(b)| = 0.$$

This result shows that all sets are of finite measure.

8.2 Let Γ be an arbitrary finite graph. We shall now investigate the independence of the inequalities (6) in the case when the values of $f(x)$, $g(x)$ are real non-negative numbers. Strictly speaking, we take as S the set of all real numbers, call measureable all subsets A of S which consist of a finite number of intervals $\alpha_\nu \leqq t < \beta_\nu$ (where $\alpha_\nu < \beta_\nu$), and we write $|A| < |A'|$ if, and only if,

$$\sum_\nu (\beta_\nu - \alpha_\nu) < \sum_\nu (\beta'_\nu - \alpha'_\nu).$$

It is obvious that in the case of non-overlapping sets $f(x)$ and $g(x)$ all operations with the sets correspond to analogous operations with their measures, and vice versa. Hence we may, for the rest of this paper, consider $f(x)$, $g(x)$ as functions defined for all points x of Γ and having as functional values real, non-negative numbers. Put

$$f(x) - g(x) = h(x).$$

Then (6) corresponds to

$$(h, \mathfrak{A}) \geqq 0 \qquad \textit{for all } \mathfrak{A}. \tag{79}$$

We want to study the interdependencies between the various inequalities (79).

Let

$$\mathfrak{A}^{(1)}, \mathfrak{A}^{(2)}, \cdots, \mathfrak{A}^{(m)}$$

be all sections of Γ,

$$\mathfrak{A}^{(\mu)} = \{x_1^{(\mu)}, x_2^{(\mu)}, \cdots, x_{n_\mu}^{(\mu)}\} \qquad (1 \leqq \mu \leqq m) \tag{80}$$

where, for every μ, the points

$$x_1^{(\mu)}, \cdots, x_{n_\mu}^{(\mu)}$$

are different from each other. The sections $\mathfrak{A}^{(\mu)}$ are also supposed to be different from each other. (79) is the same as

$$h(x_1^{(\mu)}) + h(x_2^{(\mu)}) + \cdots + h(x_{n_\mu}^{(\mu)}) \geqq 0 \qquad (1 \leqq \mu \leqq m). \tag{81}$$

Here the symbols $h(x)$ denote independent variables which can take any real values.

8.3 Define a relation

$$a \sim b$$

as having the following meaning. There exists an ordered system

$$x_0, x_1, \cdots, x_n$$

of points satisfying $x_0 = a$; $x_n = b$ which has the property that, for every $\nu(1 \leqq \nu \leqq n)$, at least one of the relations

$$x_{\nu-1} \to x_\nu; \qquad x_\nu \to x_{\nu-1}$$

holds. In particular, $x \sim x$ for every x.

A set $\mathfrak{B}$ is called *connected* if

$$a \sim b \quad \textit{whenever} \quad a, b \prec \mathfrak{B}.$$

Obviously, every set $\mathfrak{B}$ can be represented as sum

$$\mathfrak{B}_1 + \mathfrak{B}_2 + \cdots + \mathfrak{B}_k$$

of connected sets $\mathfrak{B}_\kappa$ which have mutually no points in common, and this representation is, apart from the order, unique. If $\mathfrak{B}$ is a section then every $\mathfrak{B}_\kappa$ is a section. Also

$$(h, \mathfrak{B}) = \sum_{\kappa=1}^{k} (h, \mathfrak{B}_\kappa).$$

It is therefore evident that, in (81), we need only retain those inequalities which correspond to connected sections $\mathfrak{A}^{(\mu)}$. Our next theorem asserts that the remaining inequalities are, in fact, independent from each other. Let

$$\mathfrak{A}^{(1)}, \mathfrak{A}^{(2)}, \cdots, \mathfrak{A}^{(r)}$$

be those non-empty sections which are connected. Then $1 \leqq r \leqq m$.

THEOREM 7. *The inequalities*

$$(h, \mathfrak{A}^{(\rho)}) \geqq 0 \qquad (1 \leqq \rho \leqq r) \tag{82}$$

are independent from each other, and every inequality (79) *is a consequence of* (82).

PROOF. Put, for $1 \leqq \rho \leqq r$,

$$\chi^{(\rho)}(x) = \begin{cases} 1 & (x \prec \mathfrak{A}^{(\rho)}) \\ 0 & (x \nprec \mathfrak{A}^{(\rho)}). \end{cases}$$

Then (82) can be written in the form

$$L^{(\rho)} = \sum \chi^{(\rho)}(x) h(x) \geqq 0 \qquad (1 \leqq \rho \leqq r). \tag{83}$$

The numbers $h(x)$ are to be considered as independent real variables. Let us assume that the inequalities (83) are not independent from each other. Then $r > 1$. We can number the $\mathfrak{A}^{(\rho)}$ in such a way that the inequality

$$L^{(1)} \geqq 0$$

is a consequence of the system

$$L^{(\rho)} \geqq 0 \qquad (1 < \rho \leqq r).$$

According to a theorem which follows most naturally from the theory of convex sets of points (For an elementary proof see **3**, Satz 3. A very similar theorem is proved in **4**), this can only be true if $L^{(1)}$ is a linear combination of $L^{(2)}, \cdots, L^{(r)}$ with non-negative coefficients i.e. if there exists an identity in the variables $h(x)$ which is of the form

$$\sum_x \chi^{(1)}(x)h(x) = \sum_{\rho=2}^{r} K_\rho \sum_x \chi^{(\rho)}(x)h(x)$$

where

$$K_2, \cdots, K_r \geqq 0, \text{ const.}$$

Then

$$\chi^{(1)}(x) = \sum_{\sigma=2}^{r} K_\rho \chi^{(\rho)}(x) \qquad \textit{for all } x. \tag{84}$$

$\mathfrak{A}^{(1)}$ is not empty. Let $x_1 \prec \mathfrak{A}^{(1)}$. From

$$1 = \chi^{(1)}(x_1) = \sum_{\rho=2}^{r} K_\rho \chi^{(\rho)}(x_1)$$

and

$$K_\rho \chi^{(\rho)}(x_1) \geqq 0 \qquad (1 < \rho \leqq r) \tag{85}$$

follows that at least one of the numbers (85) is positive. Hence we may assume that

$$K_2 > 0; \qquad \chi^{(2)}(x_1) = 1. \tag{86}$$

If $x \prec \mathfrak{A}^{(2)}$ then, by (84),

$$\chi^{(1)}(x) \geqq K_2 \chi^{(2)}(x) > 0,$$

$$\chi^{(1)}(x) = 1; \qquad x \prec \mathfrak{A}^{(1)}.$$

Hence $\mathfrak{A}^{(2)} \subset \mathfrak{A}^{(1)}$. Because of $\mathfrak{A}^{(2)} \neq \mathfrak{A}^{(1)}$ there exists a point

$$x_2 \prec \mathfrak{A}^{(1)} - \mathfrak{A}^{(2)}.$$

Since

$$x_1, x_2 \prec \mathfrak{A}^{(1)}$$

and $\mathfrak{A}^{(1)}$ is connected we can find a chain

$$(87) \qquad x_1 = a_0 ; \quad a_1 ; \quad \cdots ; \quad a_l = x_2$$

which is such that, for every λ satisfying $1 \leqq \lambda \leqq l$, at least one of the relations

$$a_{\lambda-1} \rightarrow a_\lambda ; \qquad a_\lambda \rightarrow a_{\lambda-1}$$

is true. Furthermore, $a_0, a_1, \cdots, a_l$ are in $\mathfrak{A}^{(1)}$. In view of

$$a_0 \prec \mathfrak{A}^{(2)}; \qquad a_l \not\prec \mathfrak{A}^{(2)}$$

there is at least one index λ_0 satisfying

$$1 \leqq \lambda_0 \leqq l; \qquad a_{\lambda_0-1} \prec \mathfrak{A}^{(2)}; \qquad a_{\lambda_0} \not\prec \mathfrak{A}^{(2)}.$$

Remembering that $\mathfrak{A}^{(2)}$ is a section we conclude that $a_{\lambda_0} \rightarrow a_{\lambda_0-1}$ is impossible, and that therefore, according to the property of the system (87), $a_{\lambda_0-1} \rightarrow a_{\lambda_0}$. Then, using again the definition of sections, we find that

$$\chi^{(\rho)}(a_{\lambda_0-1}) - \chi^{(\rho)}(a_{\lambda_0}) \begin{cases} = & 0 \ (\rho = 1) \\ = & 1 \ (\rho = 2) \\ \geqq & 0 \ (2 < \rho \leqq r). \end{cases}$$

Finally, by (84),

$$0 = \chi^{(1)}(a_{\lambda_0-1}) - \chi^{(1)}(a_{\lambda_0}) = \sum_{\rho=2}^{r} K_\rho(\chi^{(\rho)}(a_{\lambda_0-1}) - \chi^{(\rho)}(a_{\lambda_0})) \geqq K_2$$

which contradicts (86).

In conclusion I should like to add that it is possible, by means of the same theorem on linear inequalities used above, to prove the special case of Theorem 1 where T' is replaced by T and the measure is that defined in 7.2.

THE UNIVERSITY
SHEFFIELD

REFERENCES

(1) R. RADO, *A theorem on general measure functions*, Proc. Lond. Math. Soc. 44 (1938) 61–91.
(2) HARDY, LITTLEWOOD AND POLYA, Inequalities (Cambridge 1934).
(3) H. WEYL, *Elementare Theorie der konvexen Polyeder*, Comment. Math. Helvetici 7 (1935), 290–306.
(4) E. STIEMKE, *Lineare Ungleichungen*, Math. Annalen 76 (1915), 340–342.

ANNALS OF MATHEMATICS
Vol. 44, No. 2, April, 1943

FAMILY OF TOTALLY UMBILICAL HYPERSURFACES IN AN EINSTEIN SPACE

BY YUNG-CHOW WONG[1]

(Received October 5, 1942; Revised January 4, 1943 and February 10, 1943)

1. Introduction

Let V_n be a Riemannian n-space with fundamental tensor[2] $g_{ij}(h, i, j, k, l = 1, \cdots, n)$.[3] As usual we denote by a comma the covariant differentiation with respect to g_{ij}, and by R^l_{ijk} and R_{ij} the Riemann and Ricci tensors for g_{ij}, respectively. The *scalar curvature*[4] a and the tensor L_{ij} are defined by

$$(1.1) \qquad a = -\frac{R}{n(n-1)},$$

$$(1.2) \qquad L_{ij} = R_{ij} - \frac{R}{2(n-1)} g_{ij},$$

where $R = g^{ij}R_{ij}$. V_n is called an *Einstein* space, a *conformal-Euclidean* space,[5] or a space of *constant curvature*, and is denoted by E_n, C_n, or S_n, respectively, if V_n satisfies the respective condition:[6]

$$(1.3) \qquad E_n(n > 2) \qquad R_{ij} = -(n-1)ag_{ij},$$

$$(1.4) \qquad C_n(n > 2) \qquad \begin{cases} R^l_{ijk} = \dfrac{1}{n-2}(L_{i[j}\delta^l_{k]} + g_{i[j}L^l_{k]}) & \text{for} \quad n > 3, \\ L_{i[j,k]} = 0 & \text{for} \quad n = 3, \end{cases}$$

$$(1.5) \qquad S_n(n \geqq 2) \qquad R^l_{ijk} = -ag_{i[j}\delta^l_{k]}.$$

In (1.3) and in (1.5) for $n > 2$ the scalar curvature a is automatically constant, while in (1.5) for $n = 2$, a is assumed to be constant. For $n = 3$, $(1.4)_1$ is

[1] This paper was written while the author was a Chinese Ying-Keng Scholarship Student visiting Massachusetts Institute of Technology. The author wishes to thank Prof. D. J. Struik for the conversations they had from time to time during the preparation of the manuscript.

[2] Fundamental tensors are always supposed to be non-singular, though not necessarily definite. All functions appearing in this paper are real and are assumed to have differentiability properties adequate to the part they play in the discussion.

[3] An index has the same range throughout this paper. An index which appears twice in an expression, once as superscript and once as subscript, is to be summed over the appropriate range. A numerical index at the upper right-hand corner means an exponential except in the case of the coordinates x^a, x^k, x^b or x^p.

[4] Here we follow Schouten's definition; Eisenhart calls R the scalar curvature.

[5] I.e. a V_n which can be mapped conformally on a Euclidean space. In this paper whenever we speak of a C_n, it is understood that $n > 2$. For the theory of a C_n, see, e.g., Eisenhart 5, 89–92; Schouten-Struik 11, 199–205, as listed at the end of this paper.

[6] We write, e.g., $L_{ij}\delta^l_k - L_{ik}\delta^l_j = L_{i[j}\delta^l_{k]}$.

identically satisfied; for $n > 3$, $(1.4)_2$ is a consequence of $(1.4)_1$. From (1.3), (1.4) and (1.5), it follows at once that an $S_n(n > 2)$ is an E_n and a C_n; and a V_n which is an E_n and a C_n must be an S_n.

The present paper is the result of an attempt to generalize a theorem given in a previous paper of mine (Wong **14**, Th. 4.2′). Briefly stated, the theorem is: *A one-parameter family of E_n which are conformal to one another can in general be imbedded in an E_{n+1} as totally umbilical hypersurfaces.* A hypersurface in a Riemannian space is called *totally umbilical* if its first and second fundamental tensors differ by a scalar factor; in particular, it is called *totally geodesic* if its second fundamental tensor is identically zero.

In Part I of this paper we first establish a generalization of the above theorem. Then a necessary and sufficient condition is obtained for a V_n to be imbedable in an E_{n+1} as a member of ∞^1 totally umbilical hypersurfaces. The condition is that V_n have constant scalar curvature and that there exist scalar functions ρ of a certain nature satisfying the differential equation

$$\rho_{,ij} + \phi(\rho)L_{ij} \backsim g_{ij}. \tag{1.6}$$

$\phi(\rho)$ is a known polynomial in ρ, and the sign $\backsim$ indicates that the tensor on its left is equal to the tensor on its right multiplied by a scalar. The determination of the imbedding E_{n+1}, if it exists, depends solely on the solutions of (1.6). An application of a classical theorem on a system of linear differential equation reduces the compatibility of (1.6) to that of a system of polynomial equations.

A desire to determine all the C_n imbedable in an E_{n+1} as a member of ∞^1 totally umbilical hypersurfaces leads to the work of Part II, which contains a detailed study of (1.6) for the case of a C_n. For an E_n or a V_2, equation (1.6) reduces to $\rho_{,ij} \backsim g_{ij}$, which has been studied by Brinkmann (**2**), Fialkow (**8**), Yano (**13**), and Wong (**14**, §2). Here it is proved that a C_n, not an S_n, for which equation (1.6) admits a solution is characterized by its tensor L_{ij} having one of two particular forms. Canonical forms for the fundamental tensors of such C_n are obtained.

I. Totally Umbilical Hypersurfaces in an E_{n+1}

2. Family of a totally umbilical V_n in an E_{n+1}

Given in a V_m $(m = n + 1 > 2)$ a one-parameter family of hypersurfaces V_n, whose (first) fundamental tensors are of rank n, then a coordinate system exists in which the equation of V_n is $x^m =$ const. and the fundamental tensor ${}^*g_{\alpha\beta}(\alpha, \beta = 1, \cdots, m)$ is such that ${}^*g_{im} = 0 (i = 1, \cdots, n)$ (Eisenhart **5**, 144 and 5). It is known (Eisenhart **5**, 182) that in this coordinate system, V_n are totally umbilical in V_m if and only if ${}^*g_{ij}$ is of the form ${}^*g_{ij} = \rho^{-2}g_{ij}$, where $\rho = \rho(x^\alpha)$ and $g_{ij} = g_{ij}(x^k)$, and that[7] $\partial_m\rho = 0$ or $\neq 0$ according as V_n are totally geodesic or not.

[7] We write $\partial_\alpha = \dfrac{\partial}{\partial x^\alpha}$.

Let the fundamental tensor of V_m be

$$(2.1)\qquad {}^*g_{\alpha\beta} = \begin{bmatrix} \rho^{-2} g_{ij} & 0 \\ 0 & {}^*g_{mm} \end{bmatrix}.$$

The fundamental tensors of V_n are evidently $\rho^{-2}g_{ij}$. In order that they may be isometric to a given set of fundamental tensors $[{}'\rho({}'x^\alpha)]^{-2}{}'g_{ij}({}'x^k)$ with ${}'x^m$ as parameter, it is necessary and sufficient that there exists a transformation of the form

$$x^k = x^k({}'x^i), \qquad x^m = x^m({}'x^m),$$

identifying these two sets of fundamental tensors. In the coordinates ${}'x^\alpha$, the fundamental tensor of V_m is of the form (2.1) with ρ, g_{ij}, ${}^*g_{mm}$ replaced by ${}'\rho$, ${}'g_{ij}$, ${}'{}^*g_{mm}$, respectively.

The above observations furnish us a way to discuss the imbedability of a one-parameter family of conformal V_n in an E_{n+1} as totally umbilical hypersurfaces. In fact, we may suppose the tensors $\rho^{-2}g_{ij}$ in (2.1) given, and determine the condition in which ${}^*g_{mm}$ may be found so that the tensor (2.1) represents[8] an E_{n+1}. The result is the following

THEOREM 2.1. *A one-parameter family of V_n with fundamental tensors ${}^*g_{ij}(x^k, x^m)$ and parameter x^m can be imbedded in an E_{n+1} as totally umbilical hypersurfaces, which are not totally geodesic, if and only if the following conditions are satisfied:*

(1) ${}^*g_{ij}$ is of the form $[\rho(x^k, x^m)]^{-2}g_{ij}(x^k)$, where $\partial_m\rho \neq 0$;

(2) Each V_n has constant scalar curvature;

(3) When it is assumed[9] that $\rho(x^k, x_o^m) = 1$ for some value x_o^m of x^m, the tensor

$$\partial_m\left(\frac{\rho_{,ij}}{\rho^{n-1}}\right) + \frac{\partial_m\rho}{\rho^{n-1}} R_{ij}$$

differs from g_{ij} by a scalar factor.

If $n = 2$, the tensor in (3) can be replaced by $\rho_{,ij}$; if $n > 2$, it can be replaced by

$$\rho_{,ij} + \frac{\rho^{n-1} - \rho}{n-2} R_{ij} \qquad \textit{and by} \qquad {}'R_{ij} - \rho^{n-1} R_{ij},$$

where ${}'R_{ij}$ is the Ricci tensor of ${}^*g_{ij}$.

[8] By this we mean that ${}^*g_{\alpha\beta}$ is the fundamental tensor of an E_{n+1}.

[9] This assumption is not a restriction; for, if x_o^m is any fixed value of x^m,

$$[\rho(x^k, x^m)]^{-2}g_{ij}(x^k) = \left[\frac{\rho(x^k, x^m)}{\rho(x^k, x_o^m)}\right]^{-2} \{[\rho(x^k, x_o^m)]^{-2}g_{ij}(x^k)\},$$

and $\dfrac{\rho(x^k, x^m)}{\rho(x^k, x_o^m)}$, $[\rho(x^k, x_o^m)]^{-2}g_{ij}(x^k)$ can be used in place of $\rho(x^k, x^m)$, $g_{ij}(x^k)$, respectively. From this it follows that any tensor of the family $\rho^{-2}g_{ij}$ may be made to play the part of g_{ij}.

Let a one-parameter family of V_n with fundamental tensors $^*g_{ij}(x^k, x^m)$ satisfy conditions (1), (2) and (3), and let the scalar curvatures of V_n be $a(x^m)$. Then if c is any constant not identically equal to $a(x^m)$, there exists one and only one E_{n+1} whose scalar curvature is c and in which V_n can be imbedded as totally umbilical hypersurfaces; the components of the fundamental tensor $^*g_{\alpha\beta}$ of the E_{n+1} are

$$^*g_{im} = 0, \qquad ^*g_{ij} = \rho^{-2} g_{ij}, \qquad ^*g_{mm} = \frac{(\partial_m \log \rho)^2}{a(x^m) - c}.$$

PROOF. The components of the Ricci tensor $^*R_{\alpha\beta}$ of the fundamental tensor (2.1) are (Wong **14**, (3.6))

$$(2.2) \quad \begin{cases} ^*R_{im} = -(m-2)(\partial_i \rho_m + \rho_m \sigma_i), \\ ^*R_{ij} = {}'R_{ij} + \sigma\left(\dfrac{1}{\sigma}\right)_{;ij} + {}^*g_{ij}{}^*g^{mm}\left[\rho\left(\dfrac{1}{\rho}\right)_{;mm} + (m-2)\rho^2\left(\dfrac{1}{\rho}\right)_{;m}\left(\dfrac{1}{\rho}\right)_{;m}\right], \\ ^*R_{mm} = (m-1)\rho\left(\dfrac{1}{\rho}\right)_{;mm} + {}^*g_{mm}{}^*g^{ij}\sigma\left(\dfrac{1}{\sigma}\right)_{;ij}, \end{cases}$$

where σ is defined by $e\sigma^2 = {}^*g^{mm} = 1/{}^*g_{mm}$, $e = \pm 1$; $\rho_m = \partial_m \log \rho$, $\sigma_i = \partial_i \log \sigma$; $'R_{ij}$ is the Ricci tensor of $^*g_{ij} = \rho^{-2} g_{ij}$; and the semi-colon followed by the indices i and j or m denotes covariant differentiation with respect to $^*g_{ij}$ or $^*g_{mm}$, respectively.[10] The condition for $^*g_{\alpha\beta}$ to represent an E_m is

$$(2.3) \qquad ^*R_{\alpha\beta} = -(m-1)c\, {}^*g_{\alpha\beta},$$

where c is the constant scalar curvature of E_m.

We now suppose that $\partial_m \rho \neq 0$, i.e. that the totally umbilical hypersurfaces $x^m =$ const. are not totally geodesic in E_m. From (2.3) and $(2.2)_1$ we have

$$\partial_i \rho_m + \rho_m \sigma_i = 0,$$

which gives on integration

$$(2.4) \qquad \sigma\rho_m = z(x^m) \neq 0,$$

where $z(x^m)$ is some function of x^m alone. If we transvect $(2.2)_2$ and $(2.2)_3$ with $^*g^{ij}$ and $^*g^{mm}$ respectively and subtract the results, we find, on making use of (2.3) and (2.4),

$$(2.5) \qquad -\frac{^*g^{ij}\,{}'R_{ij}}{n(n-1)} = c + e[z(x^m)]^2 \equiv a(x^m).$$

Since by definition the left-hand member is the scalar curvature of $^*g_{ij}$, this equation shows that *for each value of x^m the fundamental tensor $^*g_{ij}$ has constant scalar curvature*, which we denote by $a(x^m)$. We observe that since $z(x^m) \neq 0$, $a(x^m)$ cannot be identically equal to c.

---

[10] In this operation, the x^m in $^*g_{ij}$ and the x^k in $^*g_{mm}$, respectively, are considered as parameters. The covariant derivative with respect to $^*g_{mm}$ is formed with the Christoffel symbol of the second kind $'\Gamma^m_{mm}$ for $^*g_{mm}$.

Now equations (2.4) and (2.5) can be solved for $e\sigma^{-2}$, giving

$$e\sigma^{-2} = \frac{(\rho_m)^2}{a(x^m) - c} = \frac{(\partial_m \log \rho)^2}{a(x^m) - c}. \tag{2.6}$$

Therefore the first part of our theorem will be proved if we can show that in consequence of (2.4) and (2.5), equations (2.2) and (2.3) are equivalent to condition (3) of the theorem.

Because of (2.3), equation $(2.2)_2$ is of the form

$$\left(\frac{1}{\sigma}\right)_{;ij} + \frac{1}{\sigma}\,'R_{ij} = \tau\, {}^*g_{ij}, \tag{2.7}$$

where τ is a certain scalar. The integrability condition of this equation is

$$'R^{l}_{ijk}\left(\frac{1}{\sigma}\right)_{;l} + ('R_{ij;k} - 'R_{ik;j})\frac{1}{\sigma} + 'R_{ij}\left(\frac{1}{\sigma}\right)_{;k} - 'R_{ik}\left(\frac{1}{\sigma}\right)_{;j} = g_{ij}\tau_{;k} - g_{ik}\tau_{;j},$$

where $'R^{l}_{ijk}$ is the Riemann tensor of ${}^*g_{ij}$. When we transvect this with ${}^*g^{ij}$ and then make use of the well-known identity (Eisenhart **5**, 82, (26.4))

$$'R^{i}_{k;i} = \tfrac{1}{2}\,'R_{;k}, \quad \text{where} \quad 'R = {}^*g^{ij}\,'R_{ij},$$

we find

$$\frac{1}{2\sigma}\,'R_{;k} + 'R\left(\frac{1}{\sigma}\right)_{;k} = (n-1)\tau_{;k}.$$

But ${}^*g_{ij}$ is of constant scalar curvature $a(x^m)$ and $'R = -n(n-1)a(x^m)$. Therefore the above equation becomes $-na(x^m)\left(\frac{1}{\sigma}\right)_{;k} = \tau_{;k}$, which gives, on integration,

$$\tau = -\left[n\frac{a(x^m)}{\sigma} + w(x^m)\right],$$

where $w(x^m)$ is some function of x^m alone. With this value for τ, equation (2.7) can now be written more precisely as

$$\left(\frac{1}{\sigma}\right)_{;ij} + \frac{1}{\sigma}\,'R_{ij} = -\left[n\frac{a(x^m)}{\sigma} + w(x^m)\right]{}^*g_{ij}. \tag{2.8}$$

When this expression for $\left(\frac{1}{\sigma}\right)_{;ij}$ is substituted in $(2.2)_3$, the latter becomes, because of (2.3) and (2.5),

$$e\sigma^2\rho\left(\frac{1}{\rho}\right)_{;mm} + c - a(x^m) = \sigma w(x^m). \tag{2.9}$$

Now if $'\Gamma^m_{mm}$ is the Christoffel symbol of the second kind for $^*g_{mm}$, then

$$\left(\frac{1}{\rho}\right)_{;mm} = \partial_m\left(-\frac{\rho_m}{\rho}\right) - {'\Gamma^m_{mm}}\left(-\frac{\rho_m}{\rho}\right)$$

$$= \frac{\rho_m}{\rho}\left(-\partial_m \log\frac{\rho_m}{\rho} + \tfrac{1}{2}\partial_m \log {^*g_{mm}}\right)$$

$$= \frac{\rho_m}{\rho}\,\partial_m \log\frac{\rho}{\sigma\rho_m} = \frac{\rho_m}{\rho}\left[\rho_m - \frac{z'(x^m)}{z(x^m)}\right]$$

by (2.4).[11] Substituting this in (2.9) and using (2.4) and (2.5), we have

$$w(x^m) = -ez'(x^m) = -\frac{a'(x^m)}{2z(x^m)} = -\frac{a'(x^m)}{2\sigma\rho_m}. \tag{2.10}$$

Therefore (2.8) becomes

$$\left(\frac{1}{\sigma}\right)_{;ij} + \frac{1}{\sigma}\,{'R_{ij}} = -\frac{1}{\sigma}\left[na(x^m) - \frac{a'(x^m)}{2\rho_m}\right]{^*g_{ij}}. \tag{2.11}$$

It is easily seen from the way (2.5) was derived from $(2.2)_2$ and $(2.2)_3$, that on account of (2.4) and (2.5), equations (2.2) and (2.3) are equivalent to (2.11).

We now suppose, without loss of generality, that g_{ij} is a tensor of the family $^*g_{ij}$ so that[12]

$$\rho(x^k, x^m_o) = 1 \tag{2.12}$$

for some value x^m_o of x^m. We wish to write (2.11) in terms of ρ and g_{ij}. Connecting the covariant derivatives, Ricci tensors, and scalar curvatures related to the fundamental tensors $^*g_{ij} = \rho^{-2}g_{ij}$ and g_{ij} are the following formulas (Eisenhart **5**, 89–90):

$$\left.\begin{aligned} \left(\frac{1}{\sigma}\right)_{;ij} &= \frac{1}{\rho}\left(\frac{\rho}{\sigma}\right)_{,ij} - \frac{\rho_{,ij}}{\sigma} - g^{hk}\left(\frac{1}{\sigma}\right)_{,h}\rho_{,k}g_{ij}, \\ 'R_{ij} &= R_{ij} - (n-2)\frac{\rho_{,ij}}{\rho} + g_{ij}g^{hk}\left[-\frac{\rho_{,hk}}{\rho} + (n-1)\frac{\rho_{,h}\rho_{,k}}{\rho^2}\right], \\ a(x^m) &= a(x^m_o)\rho^2 + \frac{2}{n}\rho g^{hk}\rho_{,hk} - g^{hk}\rho_{,h}\rho_{,k}, \end{aligned}\right\} \tag{2.13}$$

where as usual the comma denotes covariant derivative with respect to g_{ij}. We can now readily verify that on substitution from (2.13) and making use of (2.4), equation (2.11) becomes

$$\begin{aligned} &\frac{1}{\rho}(\partial_m\rho)_{,ij} - (n-1)\frac{\partial_m\rho}{\rho^2}\rho_{,ij} + \rho_m R_{ij} \\ &= \left[\frac{g^{hk}\partial_m(\rho_{,hk})}{n\rho} - \frac{(n-1)}{n}\frac{\partial_m\rho}{\rho^2}g^{hk}\rho_{,hk} - (n-1)a(x^m_o)\rho_m\right]g_{ij}. \end{aligned} \tag{2.14}$$

---

[11] A prime (on the right) always means differentiation.

[12] See footnote 9.

Since $(\partial_m\rho)_{,ij} = \partial_m(\rho_{,ij})$, the above equation gives an identity when transvected by g^{ij} and therefore is equivalent to

$$\frac{1}{\rho}\,\partial_m(\rho_{,ij}) - (n-1)\frac{\partial_m\rho}{\rho^2}\,\rho_{,ij} + \rho_m R_{ij} \backsim g_{ij}\,,$$

i.e.

$$\partial_m\left(\frac{\rho_{,ij}}{\rho^{n-1}}\right) + \frac{\partial_m\rho}{\rho^{n-1}}\,R_{ij} \backsim g_{ij}\,, \tag{2.15}$$

Equation (2.15) proves condition (3) of our theorem.

Suppose $n > 2$. Integrating (2.15) with respect to x^m, we have

$$\frac{\rho_{,ij}}{\rho^{n-1}} - \frac{R_{ij}}{(n-2)\rho^{n-2}} + T_{ij} \backsim g_{ij}\,, \tag{2.16}$$

where T_{ij} is an integration tensor independent of x^m. Now in consequence of (2.12), we have

$$(\rho_{,ij})_{x^m = x_0^m} = 0,$$

as follows from the very definition of partial differentiation. Therefore if we put $x^m = x_o^m$ in (2.16), the result is

$$-\frac{R_{ij}}{n-2} + T_{ij} \backsim g_{ij}\,.$$

Using this expression for T_{ij} in (2.16), we have

$$\rho_{,ij} + \frac{\rho^{n-1} - \rho}{n-2}\,R_{ij} \backsim g_{ij}\,, \tag{2.17}$$

which is equivalent to (2.15). Equation (2.15) is also equivalent to

$$'R_{ij} - \rho^{n-2}R_{ij} \sim g_{ij}\,, \tag{2.18}$$

as follows by elimination of $\rho_{,ij}$ from $(2.13)_2$ and (2.17).

Finally, if $n = 2$, the equation $R_{ij} \sim g_{ij}$ is satisfied, and therefore (2.15) becomes

$$\partial_m\left(\frac{\rho_{,ij}}{\rho}\right) \backsim g_{ij}\,.$$

If we integrate this with respect to x^m and treat the result as we did (2.16), we can show that (2.15) is equivalent to

$$\rho_{,ij} \sim g_{ij}\,. \tag{2.19}$$

Equations (2.17), (2.18) and (2.19) complete the proof of our theorem.

We end this section by obtaining a direct consequence of Theorem 2.1. Let $n > 2$. If $R_{ij} = -(n-1)a(x_o^m)g_{ij}$, then (2.18) demands that $'R_{ij} = -(n-1)a(x^m)^*g_{ij}$. Also, if ρ is a function of x^m alone so that $'R_{ij} = R_{ij}$,

equation (2.18) becomes $'R_{ij} = -(n-1)a(x^m)^*g_{ij}$. Hence if we remember that any tensor of the family $^*g_{ij}$ may be made to play the part of g_{ij}, we have

THEOREM 2.2. *Let an $E_{n+1}(n > 2)$ admit a one-parameter family of totally umbilical, but not totally geodesic, hypersurfaces V_n. Then the V_n are all E_n, if one of them is an E_n or if every V_n is trivially conformal*[13] *to another. Conversely, let c be a constant and $[\rho(x^k, x^m)]^{-2}g_{ij}(x^k)$ with x^m as parameter be the fundamental tensors of ∞^1 conformal $E_n(n > 2)$ whose scalar curvatures are not all equal to c. Then if $\partial_m\rho \neq 0$, there exists a unique E_{n+1} of scalar curvature c in which the given E_n can be imbedded as totally umbilical hypersurfaces.*

The latter part of this theorem has been mentioned at the beginning of this paper.

3.1 An imbedding problem

Suppose that for a V_n the equation

$$(3.1)\qquad \begin{cases} \rho_{,ij} + \dfrac{\rho^{n-1} - \rho}{n-2} R_{ij} \backsim g_{ij} & \text{for } n > 2, \\ \rho_{,ij} \backsim g_{ij} & \text{for } n = 2 \end{cases}$$

admits a solution for ρ. We now prove that if the scalar curvature of V_n is constant, then that of $\rho^{-2}g_{ij}$ is also constant. This result combined with Theorem 2.1 gives the following

THEOREM 3.1. *A necessary and sufficient condition for a V_n with fundamental tensor g_{ij} to be imbedable in an E_{n+1} as a member of ∞^1 totally umbilical, but not totally geodesic, hypersurfaces is that V_n has constant scalar curvature and equation (3.1) admits a one-parameter family of solutions $\rho(x^k, x^m)$ which is such that $\rho(x^k, x_0^m) = 1$ for some value x_0^m of the parameter x^m. If a V_n satisfies this condition and ρ is any one-parameter family of solutions of (3.1) of the above-mentioned nature, then $\rho^{-2}g_{ij}$ are fundamental tensors of a family of V_n which can be imbedded in an E_{n+1} as totally umbilical hypersurfaces.*

PROOF. Suppose that the scalar curvature $a = -(R/n(n-1))$ of V_n is constant. Consider first the case $n > 2$. We write $(3.1)_1$ as

$$(3.1')_1\qquad \rho_{,ij} + \phi R_{ij} = \psi g_{ij},$$

where ψ is an unknown scalar and

$$(3.2)\qquad \phi = \frac{\rho^{n-1} - \rho}{n-2}.$$

The integrability condition of $(3.1')_1$ is

$$R_{i\,jk}^{l}\rho_{,l} + \phi' R_{i[j}\rho_{,k]} + \phi R_{i[j,k]} = g_{i[j}\psi_{,k]}.$$

[13] Two V_n are said to be *trivially conformal* to each other, if their fundamental tensors (in the same coordinates) differ by a constant scalar factor.

Transvection of this with g^{ij} gives, because $R_{i[j,k]}g^{ij} = \frac{1}{2}R_{,k} = 0$ (Eisenhart **5**, 82, (26.4)),

$$R^l_k\rho_{,l} = \frac{n-1}{1-\phi'}(\psi_{,k} + na\phi'\rho_{,k}). \tag{3.3}$$

Using $(3.1')_1$, (3.2) and (3.3), we have

$$\begin{aligned}(g^{hk}\rho_{,h}\rho_{,k})_{,i} &= 2g^{hk}\rho_{,h}\rho_{,ki}\\ &= 2[(\rho\psi)_{,i} + na\phi'\rho\rho_{,i}].\end{aligned}$$

Let us then substitute for ϕ' from (3.2) and integrate; this gives

$$g^{hk}\rho_{,h}\rho_{,k} = 2\rho\psi + \frac{2(n-1)a\rho^n - na\rho^2}{n-2} - A, \qquad A = \text{const.} \tag{3.4}$$

Now the scalar curvature $'a$ of $\rho^{-2}g_{ij}$ is (cf. (2.13))

$$'a = a\rho^2 + \tfrac{1}{2}n\rho g^{hk}\rho_{,hk} - g^{hk}\rho_{,h}\rho_{,k},$$

and it is readily shown that on account of $(3.1')_1$, (3.2) and (3.4), the right-hand member of the above equation is equal to A. Hence the scalar curvature of $\rho^{-2}g_{ij}$ is constant, and our theorem for the case $n > 2$ is proved.

Equation $(3.1)_2$ for the case $n = 2$ can be treated in like manner. Indeed, we have in this case

$$\rho_{,ij} = \psi g_{ij}, \qquad R^l_k\rho_{,l} = \psi_{,k},$$
$$R_{hk} = -ag_{hk}, \qquad \psi = -(a\rho + B),$$

where B is a constant. The first two of these equations correspond to $(3.1')_1$ and (3.3) respectively; the third holds because $n = 2$; and the last follows at once from the second and third equations. We can then easily verify that

$$g^{hk}\rho_{,h}\rho_{,k} = -(a\rho^2 + 2B\rho + A), \qquad A = \text{const.},$$

and hence that A is the scalar curvature of $\rho^{-2}g_{ij}$. Our theorem is thus completely proved.

3.2. Continuation

The question naturally arises: Given a V_n with constant scalar curvature, how can we know whether or not it is imbedable in an E_{n+1} as a member of ∞^1 totally umbilical hypersurfaces? The answer given in Theorem 3.1 is not explicit enough, but may be so reduced that it depends only on the compatibility of a system of polynomial equations.

Consider first the general case $n > 2$. It is readily seen from $(3.1')_1$, (3.2) and (3.3) that equation $(3.1)_1$ is equivalent to the following system of linear differential equations in the $n + 2$ unknown functions ρ, ψ, θ_i:

$$\begin{cases}\theta_{i,j} + \phi(\rho)R_{ij} = \psi g_{ij},\\[2mm] \psi_{,i} = \dfrac{1-\phi'}{n-1}R^l_i\theta_l - na\phi'\theta_i,\\[2mm] \rho_{,i} = \theta_i.\end{cases} \tag{3.5}$$

We now apply a classical theorem (Eisenhart **6**, 1–4) on a system of linear differential equations to (3.5). The integrability conditions of $(3.5)_1$, $(3.5)_2$, $(3.5)_3$ are respectively

$$(3.6)\quad \begin{cases} R^l_{i\,jk}\theta_l + \phi' R_{i[\,j}\rho_{,k]} + \phi R_{i[\,j,k]} = g_{i[\,j}\psi_{,k]}\,, \\ (1-\phi')\theta_l R^l_{[i,j]} = (1-\phi')\theta_l R^l_{[i,j]} + (1-\phi')\theta_{l,[\,j}R^l_{i]} - \phi''\theta_l R^l_{[i}\rho_{,j]} = 0, \\ \theta_{[i,j]} = 0. \end{cases}$$

Equation $(3.6)_3$ is satisfied on account of $(3.5)_1$. On substitution from (3.5), equations $(3.6)_1$ and $(3.6)_2$ become

$$(3.7)\quad \begin{cases} \theta_l\left[R^l_{ijk} - \dfrac{1}{n-1}\, g_{i[j}R^l_{k]} + \phi'\left(R_{i[j}\delta^l_{k]} + \dfrac{1}{n-1}\, g_{i[j}R^l_{k]} + na\, g_{i[j}\delta^l_{k]}\right)\right] \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad + \phi R_{i[j,k]} = 0, \\ (1-\phi')\theta_l R^l_{[i,j]} - \phi''\,\theta_l\theta_h R^l_{[i}\delta^h_{j]} = 0. \end{cases}$$

It is easily verified that $(3.7)_2$ is identical with $(3.7)_1$ transvected by θ^i. This was also to be expected, because (3.3) was obtained from the integrability condition of $(3.1')_1$. Hence the integrability condition of (3.5) reduces to $(3.7)_1$, which we write as

$$(F)_1 \qquad (A^l_{i\,jk} + B^l_{i\,jk}\phi')\theta_l + C_{i\,jk}\phi = 0,$$

where

$$(3.8)\quad \begin{cases} A^l_{ijk} = R^l_{ijk} - \dfrac{1}{n-1}\, g_{i[j}R^l_{k]}\,, \\ B^l_{ijk} = R_{i[j}\delta^l_{k]} + \dfrac{1}{n-1}\, g_{i[j}R^l_{k]} + na g_{i[j}\delta^l_{k]}\,; \\ C_{ijk} = R_{i[j,k]} \end{cases}$$

are tensors constructed from g_{ij}.

It is not difficult to show that $(F)_1$ is identically satisfied if and only if V_n is an S_n. If $(F)_1$ is not identically satisfied, we differentiate it covariantly with respect to g_{ij} and substitute for $\theta_{i,j}$, $\psi_{,i}$ and $\rho_{,i}$ from (3.5). The result is

$$(F)_2\quad \begin{aligned} &\phi'' B^l_{i\,jk}\theta_l\theta_h + [A^l_{i\,jk,h} + (B^l_{i\,jk,h} + C_{i\,jk}\delta^l_h)\phi']\theta_l \\ &\qquad\qquad + (A^l_{i\,jk} + B^l_{i\,jk}\phi')(-\phi R_{lh} + \psi g_{lh}) + C_{i\,jk,h}\phi = 0. \end{aligned}$$

Because of $(F)_1$, this cannot be identically satisfied without $(F)_1$ first being identically satisfied. Differentiating $(F)_2$ covariantly and then substituting for $\theta_{i,j}$, $\psi_{,i}$ and $\rho_{,i}$ from (3.5), we get a set of equations $(F)_3$. Proceeding in this way we obtain a sequence of sets $(F)_1$, $(F)_2$, etc. of equations, to which a classical theorem on mixed systems of total differential equations (Eisenhart **6**, 3) can be directly applied. We observe that the equations $(F)_1$, $(F)_2$, etc. in our

case are all polynomial equations and that they can be constructed as soon as the fundamental tensor g_{ij} of V_n is given.

Hence combining the preceding results with Theorem 3.1 we have the necessity of the conditions in the following

THEOREM 3.2. *In order that a $V_n (n > 2)$ may be imbedded in an E_{n+1} as a member of ∞^1 totally umbilical, but not totally geodesic, hypersurfaces, it is necessary and sufficient that the following conditions are satisfied:*

(1) A positive integer N exists such that the equations of the sets $(F)_1$, $(F)_2$, $\cdots$, $(F)_N$ admit a one-parameter family of solutions $\rho(x^k, x^m)$, $\psi(x^k, x^m)$, $\theta_i(x^k, x^m)$ of the nature that $\rho(x^k, x_o^m) = 1$ for some value x_o^m of the parameter x^m;

(2) The equations of the set $(F)_{N+1}$ are satisfied because of the equations of the sets $(F)_1$, $(F)_2$, $\cdots$, $(F)_N$.

To prove the sufficiency of the conditions we need only show that a V_n satisfying them must be of constant scalar curvature. But the latter fact is easily proved; for, if we transvect $(F)_1$ with g^{ij}, the result is $\frac{1}{2}\phi R_{,k} = 0$, showing that R is constant.

A V_2 with constant scalar curvature is an S_2. For an S_n equation (3.1) becomes

$$\rho_{,ij} \backsim g_{ij}. \tag{3.9}$$

If we use the Riemann form of the fundamental form of S_n (Eisenhart **5**, 85)

$$g_{ij}\, dx^i\, dx^j = \frac{e_1\,(dx^1)^2 + \cdots + e_n\,(dx^n)^2}{\left\{1 + \frac{a}{4}\,[e_1(x^1)^2 + \cdots + e_n(x^n)^2]\right\}^2} \qquad (e_i = \pm 1), \tag{3.10}$$

where a is the (constant) scalar curvature of S_n, then the general solution of (3.9) is readily seen to be

$$\rho = \frac{A_o[e_1(x^1)^2 + \cdots + e_n(x^n)^2] + A_i x^i + A_{n+1}}{1 + \frac{a}{4}\,[e_1(x^1)^2 + \cdots + e_n(x^n)^2]}, \tag{3.11}$$

where the A's are arbitrary constants Evidently, by suitable choice of the A's as functions of a parameter x^m, a one-parameter family of solutions $\rho(x^k, x^m)$ can be found (and in infinitely many ways) such that $\rho(x^k, x_o^m) = 1$ for some value x_o^m of x^m. Hence the condition of Theorem 3.1 is satisfied. Now it is seen from (3.10) and (3.11) that in this case every tensor of the family $\rho^{-2}g_{ij}$ is the fundamental tensor of an S_n. And since it is known (Wong **14**, Th. 9.3) that an E_{n+1} which admits $\infty^1 S_n$ as totally umbilical hypersurfaces is an S_{n+1}, we have

THEOREM 3.3. *An S_n can always be imbedded (and in infinitely many ways) in an S_{n+1} as a member of ∞^1 totally umbilical hypersurfaces.*

4. Totally geodesic hypersurfaces in an E_{n+1}

We now consider the case excluded from the preceding discussions. As follows from what we said at the beginning of §2, the condition for a V_n with

fundamental tensor g_{ij} to be imbedable in an E_m $(m = n + 1)$ as a member of ∞^1 totally geodesic hypersurfaces is equivalent to the condition for the existance of a non-zero scalar $\rho(x^\alpha)$ such that the fundamental tensor $^*g_{\alpha\beta}$ with components

$$(4.1) \qquad ^*g_{im} = 0, \qquad ^*g_{ij} = g_{ij}(x^k), \qquad ^*g_{mm} = e\rho^2, \qquad (e = \pm 1),$$

represents an E_m. By definition, $^*g_{\alpha\beta}$ represents an E_m if and only if

$$(4.2) \qquad ^*R_{\alpha\beta} = -nc\, ^*g_{\alpha\beta}, \qquad c = \text{const.}$$

Now the components of the Ricci tensor of $^*g_{\alpha\beta}$ are

$$(4.3) \qquad ^*R_{im} = 0, \qquad ^*R_{ij} = R_{ij} + \frac{\rho_{,ij}}{\rho}, \qquad ^*R_{mm} = {}^*g_{mm}g^{ij}\frac{\rho_{,ij}}{\rho}.$$

Therefore (4.2) becomes, on account of (4.1),

$$(4.4) \qquad \rho_{,ij} + \rho R_{ij} = -nc\rho g_{ij}, \qquad g^{ij}\rho_{,ij} = -nc\rho.$$

Transvecting $(4.4)_1$ with g^{ij} and comparing the result with $(4.4)_2$, we find

$$(4.5) \qquad R = g^{ij}R_{ij} = -n(n-1)c.$$

This shows that the scalar curvature a of V_n is equal to the (constant) scalar curvature c of E_m. Hence from $(4.4)_1$ we have the following theorem complementary to Theorem 3.1 of the preceding section:

THEOREM 4.1. *A necessary and sufficient condition for a V_n with fundamental tensor g_{ij} to be imbedable in an E_{n+1} as a member of ∞^1 totally geodesic hypersurfaces is that the scalar curvature a of V_n is constant and the equation*

$$(4.6) \qquad \rho_{,ij} + \rho R_{ij} = -na\rho g_{ij}$$

admits a non-zero solution for ρ. If a V_n satisfies this condition, then corresponding to each non-zero solution ρ of (4.6) there are two and only two such imbedding E_{n+1}, and the components of their fundamental tensors $^*g_{\alpha\beta}$ are given by (4.1).

Equation (4.6) is equivalent to the following system of linear differential equations in the $n + 1$ unknown functions ρ and θ_i:

$$(4.7) \qquad \theta_{i,j} + \rho R_{ij} = -na\rho g_{ij}, \qquad \rho_{,i} = \theta_i,$$

for which the sets $(G)_1$, $(G)_2$, etc. of equations corresponding to $(F)_1$, $(F)_2$, etc. of §3.2 are

$$(G_1) \qquad (A^l_{ijk} + B^l_{ijk})\theta_l + C_{ijk}\rho = 0,$$

$$(G_2) \quad (A^l_{ijk,h} + B^l_{ijk,h} + C_{ijk}\delta^l_h)\theta_l - [(A^l_{ijk} + B^l_{ijk})(R_{lh} + nag_{lh}) - C_{ijk,h}]\rho = 0,$$

etc.,

where the tensors A, B and C are as defined in (3.8). Since the equations in $(G)_1$, $(G)_2$, etc. are all linear homogeneous in ρ and θ_i, we have from Theorem 4.1 the following theorem corresponding to Theorem 3.2:

THEOREM 4.2. *In order that a V_n may be imbedded in an E_{n+1} as a member of ∞^1 totally geodesic hypersurfaces, it is necessary and sufficient that the following conditions are satisfied:*

(1) A positive integer N exists such that the sets of equations $(G)_1, (G)_2, \cdots, (G)_N$ are compatible and the equations of the sets $(G)_{N+1}$ are satisfied because of the former sets;

(2) The sets $(G)_1, (G)_2, \cdots, (G)_N$ have at most n independent equations.

For an S_n, equations (4.7) become

$$\theta_{i,j} = -a\rho g_{ij}, \qquad \rho_{,i} = \theta_i, \tag{4.8}$$

which are completely integrable. Hence by the remark just above Theorem 3.3, we have

THEOREM 4.3. *An S_n can always be imbedded (and in infinitely many ways) in an S_{n+1} as a member of ∞^1 totally geodesic hypersurfaces.*

The case of an S_n is closed by Theorems 3.3 and 4.3, and will therefore be excluded from our future discussions.

5. Totally umbilical hypersurfaces in an E_{n+1} which are C_n

Equations (3.1) and (4.6) which appear in Theorems 3.1 and 4.1 are all particular cases of equation (6.1) of the next section, where the latter equation will be studied in detail for the case of a C_n, i.e. a conformal-Euclidean V_n. In light of what we shall obtain, the following theorems are immediate consequences of Theorems 3.1 and 4.1:

THEOREM 5.1. *If an $E_{n+1}(n > 2)$ admits a family of totally umbilical hypersurfaces which are C_n but not S_n, then the tensor L_{ij} of each of these hypersurfaces is of the form (6.3) or (6.4). Conversely, a C_n of constant scalar curvature with tensor L_{ij} of the form (6.3) or (6.4) can be imbedded in infinitely many E_{n+1} as a member of ∞^1 totally umbilical hypersurfaces or totally geodesic hypersurfaces. The determination of all these E_{n+1} is based either on a simple integration or on the solution of a linear differential equation of the second order.* (Cf. Ths. 6.1, 7.1–7.3)

THEOREM 5.2. *A C_n can be imbedded in an E_{n+1} as a member of ∞^1 totally umbilical hypersurfaces or totally geodesic hypersurfaces, if and only if it is an S_n or if $n > 2$ and its fundamental form can be reduced to the form* (8.8′), (8.9′) *or* (8.27). (Cf. Ths. 8.1, 8.2.)

II. SOME CONFORMAL-EUCLIDEAN SPACES C_n

6.1. A differential equation

For the rest of this paper we confine ourselves to the case of a $C_n(n > 2)$, and study in detail the equation

$$\rho_{,ij} + \phi L_{ij} = \omega g_{ij}, \tag{6.1}$$

where L_{ij} is the tensor (1.2), ω is an unspecified scalar, and $\phi = \phi(\rho)$ is a given function of ρ such that $\phi' \neq -1/(n-2)$.

The supposition $\phi' \neq -1/(n-2)$ is significant, because for any C_n the equation

$$(6.2) \qquad \rho_{,ij} - \frac{\rho + A}{n-2} L_{ij} = \omega g_{ij}, \qquad A = \text{const.}$$

always has a solution. Indeed, if g_{ij} represents a C_n, a scalar σ exists such that the tensor $\sigma^{-2} g_{ij}$ represents a Euclidean space. Consequently σ satisfies the equation

$$R_{ij} - (n-2)\frac{\sigma_{,ij}}{\sigma} + g_{ij}\left[-\frac{g^{hk}\sigma_{,hk}}{\sigma} + (n-1)\frac{g^{hk}\sigma_{,h}\sigma_{,k}}{\sigma^2}\right] = 0,$$

the left-hand member being the Ricci tensor of $\sigma^{-2}g_{ij}$ written in terms of g_{ij} (cf. $(2.13)_2$). From this it follows that $\rho = \sigma - A$ is a solution of (6.2).

We now proceed to prove the following

Theorem 6.1. *For a C_n, which is not an S_n, equation* (6.1) *has a solution for ρ, if and only if the tensor L_{ij} of C_n is of the form*[14]

$$(6.3) \qquad L_{ij} = \eta(\bar\rho)\bar\rho_{,i}\bar\rho_{,j} + \zeta(\bar\rho)g_{ij}, \qquad g_{ij}\bar\rho_{,i}\bar\rho_{,j} \neq 0,$$

or

$$(6.4) \qquad L_{ij} = \eta(\bar\rho)\bar\rho_{,i}\bar\rho_{,j}, \qquad g^{ij}\bar\rho_{,i}\bar\rho_{,j} = 0,$$

where $\eta(\bar\rho) \neq 0$, $\zeta(\bar\rho) \neq 0$; in the latter case $L = g^{ij}L_{ij} = 0$, $\omega = 0$. Let a C_n satisfy either of these conditions. Then a function $\bar\lambda = \bar\lambda(\bar\rho)$ exists such that

$$(6.5) \qquad \bar\rho_{,ij} = \bar\lambda\bar\rho_{,i}\bar\rho_{,j} + \frac{\zeta'}{\eta} g_{ij}, \qquad \text{or} \qquad \bar\rho_{,ij} = \bar\lambda\bar\rho_{,i}\bar\rho_{,j},$$

according as we have (6.3) *or* (6.4); *and the solution of* (6.1) *is identical with the solution $\rho = \rho(\bar\rho)$ of the linear differential equation*

$$(6.6) \qquad \rho'' + \bar\lambda\rho' + \phi\eta = 0.$$

Proof. For a C_n, equations (1.4) hold. The integrability condition of (6.1) is

$$R^l_{ijk}\rho_{,l} + \phi' L_{i[j}\rho_{,k]} + \phi L_{i[j,k]} = g_{i[j}\omega_{,k]},$$

which, on account of (1.4), becomes

$$\frac{1}{n-2}(L_{i[j}\rho_{,k]} + g_{i[j}L^l_{k]}\rho_{,l}) + \phi' L_{i[j}\rho_{,k]} = g_{i[j}\omega_{,k]},$$

i.e.

$$(6.7) \qquad \left(\phi' + \frac{1}{n-2}\right) L_{i[j}\rho_{,k]} + g_{i[j}\left(\frac{1}{n-2}L^l_{k]}\rho_{,l} - \omega_{,k]}\right) = 0.$$

---

[14] That V_n be a C_n satisfying (6.3) is, incidentally, the condition for a V_n $(n > 2)$, not an S_n, to be a *subprojective space* in the sense of Kagan. Cf. Schouten-Struik **11**, 215–225, and also the literatures quoted there.

Transvection of this with g^{ij} gives

$$(6.8) \qquad (1 - \phi')L^l_k \rho_{,l} - (n - 1)\omega_{,k} = -\left(\phi' + \frac{1}{n-2}\right) L\rho_{,k}\,.$$

Transvecting (6.7) again with $g^{hi}\rho_{,h}$, we find

$$\phi'\rho_{,l}L^l_{[j}\rho_{,k]} + \omega_{,[j}\rho_{,k]} = 0,$$

which is equivalent to

$$(6.9) \qquad \phi' L^l_k \rho_{,l} + \omega_{,k} = \alpha\rho_{,k}\,, \qquad \alpha = \text{scalar}.$$

Since $\phi' \neq -1/(n - 2)$, equations (6.8) and (6.9) can be solved for $L^l_k\rho_{,l}$ and $\omega_{,k}$, giving

$$(6.10) \qquad L^l_k\rho_{,l} = \beta\rho_{,k}\,, \qquad \omega_{,k} = \gamma\rho_{,k}\,,$$

where β and γ are some scalars. Equation $(6.10)_2$ shows that $\omega = \omega(\rho)$ and $\gamma = \omega'$. Consequently, as a result of (6.10), the integrability condition (6.7) becomes

$$(6.11) \qquad \left(\phi' + \frac{1}{n-2}\right) L_{i[j}\rho_{,k]} + \left(\frac{1}{n-2}\beta - \omega'\right) g_{i[j}\rho_{,k]} = 0,$$

which can be written in the form

$$(6.12) \qquad L_{ij} = \eta\rho_{,i}\rho_{,j} + \zeta g_{ij}\,.$$

We shall now proceed to prove that *η and ζ are both functions of ρ alone.*

On account of (6.12) we have

$$(6.13) \qquad L = \xi\eta + n\zeta,$$

$$(6.14) \qquad [2 + (n - 2)\phi']\zeta = (n - 2)\omega' - \xi\eta,$$

where

$$(6.15) \qquad \xi \equiv g^{ij}\rho_{,i}\rho_{,j}\,.$$

Equation (6.13) is the result of transvecting (6.12) with g^{ij}. To prove (6.14), we use (6.12) in (6.8) and obtain

$$(1 - \phi')(\xi\eta + \zeta) - (n - 1)\omega' = -\left(\phi' + \frac{1}{n-2}\right) L,$$

which gives (6.14) on substitution of L from (6.13).

Because of (6.12), equation (6.1) becomes

$$(6.16) \qquad \rho_{,ij} = -\phi\eta\rho_{,i}\rho_{,j} + (\omega - \phi\zeta)g_{ij}\,.$$

Thus, by differentiating (6.15) we have

$$\begin{aligned}\xi_{,k} &= 2g^{ij}\rho_{,i}\rho_{,jk}\\ &= -2(\phi\xi\eta + \phi\zeta - \omega)\rho_{,k}\,.\end{aligned}$$

From this it follows that $\xi = \xi(\rho)$ and

$$(\xi\eta + \zeta)\phi - \omega = -\tfrac{1}{2}\xi'. \tag{6.17}$$

Using (6.12) in $(1.4)_2$, we have

$$g_{i[j}\zeta_{,k]} + \eta\rho_{,i[k}\rho_{,j]} + \rho_{,i}\rho_{,[j}\eta_{,k]} = 0.$$

On substitution from (6.16), this becomes

$$g_{i[j}\zeta_{,k]} - \eta(\omega - \phi\zeta)g_{i[j}\rho_{,k]} + \rho_{,i}\rho_{,[j}\eta_{,k]} = 0. \tag{6.18}$$

At this stage, we have to consider separately the two cases $\xi \neq 0$ and $\xi = 0$.

6.2. The case $\xi \neq 0$

Since ϕ and ω are functions of ρ, it follows either from (6.14) or from (6.17) that ζ is of the form

$$\zeta = \mu(\rho) + \nu(\rho)\eta, \tag{6.19}$$

where μ and ν are some functions of ρ, and $\nu \neq 0$ because $\xi \neq 0$. Using this value of ζ in (6.18), the latter becomes

$$(\nu g_{i[j} + \rho_{,i}\rho_{,[j})\eta_{,k]} + [\mu' + \nu'\eta - \eta(\omega - \phi\zeta)]g_{i[j}\rho_{,k]} = 0.$$

Since $\nu \neq 0$, this equation can be written

$$(\nu g_{i[j} + \rho_{,i}\rho_{,[j})(\eta_{,k]} - \tau\rho_{,k]}) = 0, \tag{6.20}$$

where

$$-\nu\tau = \mu' + \nu'\eta - \eta(\omega - \phi\zeta). \tag{6.21}$$

From (6.20) it follows that either

$$\eta_{,k} - \tau\rho_{,k} = 0, \tag{6.22}$$

or $$\nu g_{ij} + \rho_{,i}\rho_{,j} = \epsilon(\eta_{,i} - \tau\rho_{,i})(\eta_{,j} - \tau\rho_{,j}), \qquad \epsilon = \text{a scalar}.$$

But the latter case cannot happen; otherwise g_{ij} would be of rank $\leqq 2$. Therefore (6.22) is true, and η is a function of ρ alone. Then it follows from (6.13) and (6.14) that ζ and L are also functions of ρ alone, as was to be proved.

Since τ is now equal to η', we have from (6.21) and (6.19) that

$$\zeta' = \eta(\omega - \phi\zeta). \tag{6.23}$$

If $\eta = 0$, it follows from (6.12) and (1.4) that C_n is an S_n. And from (6.14), (6.17), (6.23) and $\xi \neq 0$, we can easily show that if $\zeta = 0$, then $\eta = 0$ and consequently C_n is an S_n.

We have now proved the necessity of condition (6.3) for the case $\xi \neq 0$. That this same condition is also sufficient can be proved as follows.

Let a C_n be given satisfying (6.3). Then C_n cannot be an S_n. Using (6.3) in $(1.4)_2$, which is true for any C_n, we find

$$\bar{\eta}\bar{\rho}_{,i[k}\bar{\rho}_{,j]} + \bar{\zeta}'g_{i[j}\bar{\rho}_{,k]} = 0.$$

Because $\bar{\eta} \neq 0$, this can be written

$$\bar{\rho}_{,ij} = \bar{\lambda}\bar{\rho}_{,i}\bar{\rho}_{,j} + \frac{\bar{\zeta}'}{\bar{\eta}} g_{ij}, \qquad \bar{\lambda} = \text{a scalar.} \tag{6.24}$$

Differentiating (6.15) and making use of (6.24), we have

$$\bar{\xi}_{,k} = 2g^{ij}\bar{\rho}_{,i}\bar{\rho}_{,jk} = 2g^{ij}\bar{\rho}_{,i}\left(\bar{\lambda}\bar{\rho}_{,j}\bar{\rho}_{,k} + \frac{\bar{\zeta}'}{\bar{\eta}} g_{jk}\right) = 2\left(\bar{\lambda}\bar{\xi} + \frac{\bar{\zeta}'}{\bar{\eta}}\right)\bar{\rho}_{,k}.$$

Since $\bar{\xi} \neq 0$, it follows from this that $\bar{\xi}$ and $\bar{\lambda}$ are functions of $\bar{\rho}$ alone and

$$\bar{\xi} = 2\left(\bar{\lambda}\bar{\xi} + \frac{\bar{\zeta}'}{\bar{\eta}}\right). \tag{6.25}$$

If C_n, which is not an S_n, permits a solution of (6.1) such that $\xi = g^{ij}\rho_{,i}\rho_{,j} \neq 0$, equations (6.12) and (6.16) with $\eta \neq 0$, $\zeta \neq 0$ must hold. Comparison of $(6.3)_1$ with (6.12) tells us that $\rho_{,i}$ and $\bar{\rho}_{,i}$ have the same direction, which is a principal direction of L_{ij}. Hence $\bar{\rho} = \bar{\rho}(\rho)$, the exact form of which will presently be determined.

When $\bar{\rho} = \bar{\rho}(\rho)$ is used in $(6.3)_1$ and (6.24), the result is

$$L_{ij} = \bar{\eta}\bar{\rho}'^2\rho_{,i}\rho_{,j} + \bar{\zeta} g_{ij}, \tag{6.26}$$

$$\bar{\rho}'\rho_{,ij} = (\bar{\lambda}\bar{\rho}'^2 - \bar{\rho}'')\rho_{,i}\rho_{,j} + \frac{\bar{\zeta}'}{\bar{\eta}} g_{ij}. \tag{6.27}$$

Comparing these with (6.12) and (6.16) respectively, we find

$$\eta = \bar{\eta}\bar{\rho}'^2, \qquad \zeta = \bar{\zeta}, \qquad \bar{\rho}' = \frac{\bar{\lambda}\bar{\rho}'^2 - \bar{\rho}''}{-\phi\eta} = \frac{\bar{\zeta}'/\bar{\eta}}{\omega - \phi\zeta}.$$

From these it follows, by elimination of η and ζ, that

$$\phi = \frac{\bar{\rho}'' - \bar{\lambda}\bar{\rho}'^2}{\bar{\eta}\bar{\rho}'^3}, \tag{6.28}$$

$$\omega = \bar{\zeta}\phi + \frac{\bar{\zeta}'}{\bar{\eta}\bar{\rho}'} = \frac{(\bar{\rho}'' - \bar{\lambda}\bar{\rho}'^2)\bar{\zeta} + \bar{\zeta}'\bar{\rho}'^2}{\bar{\eta}\bar{\rho}'^3}. \tag{6.29}$$

Equation (6.28) is a relation between $\bar{\rho}$ and ρ, while (6.29) gives the value of the unspecified scalar ω. These equations must be satisfied if C_n permits a solution ρ of (6.1) such that $\xi \neq 0$. Now because of (6.28) and (6.29), equation (6.1) can be written

$$\bar{\eta}\bar{\rho}'^3\rho_{,ij} + (\bar{\rho}'' - \bar{\lambda}\bar{\rho}'^2)L_{ij} = [(\bar{\rho}'' - \bar{\lambda}\bar{\rho}'^2)\bar{\zeta} + \bar{\zeta}'\bar{\rho}'^2]g_{ij}.$$

But it will be noticed that this equation is identical with the result of eliminating $\rho_{,i}\rho_{,j}$ from (6.26) and (6.27) Hence, equation (6.1) is a consequence of (6.28), and we may conclude that for a C_n satisfying (6.3), the solution of (6.1) is identical with the inverted solution $\rho = \rho(\bar{\rho})$ of (6.28). But it is readily verified that written with ρ as dependent variable, (6.28) becomes (6.6). Hence our theorem for the case $\xi \neq 0$ is proved.

6.3. The case $\xi = 0$

In the preceding discussion the supposition $\xi \neq 0$ was first introduced after equation (6.18). Therefore this equation and the results preceding it also hold for the present case.

From (6.13), (6.14) and (6.17) we have, by putting $\xi = 0$,

$$(6.30) \qquad L = n\zeta, \qquad [2 + (n-2)\phi']\zeta = (n-2)\omega' \qquad \omega = \phi\zeta.$$

Since ϕ and ω are known to be functions of ρ alone, it follows from the above equations that *ζ and L are also functions of ρ alone.* As a consequence of this and (6.30), equation (6.18) now becomes

$$(6.31) \qquad (\zeta' g_{i[j} - \rho_{,i}\eta_{,[j})\rho_{,k]} = 0,$$

i.e.

$$(6.31') \qquad \zeta' g_{ij} = \rho_{,i}\eta_{,j} + \alpha_i\rho_{,j} = \rho_{,j}\eta_{,i} + \alpha_j\rho_{,i},$$

where α_i is a certain vector. The last equation shows that

$$\alpha_i = \eta_{,i} + \beta\rho_{,i}, \qquad \beta = \text{a scalar},$$

and therefore (6.31') becomes

$$\zeta' g_{ij} = \rho_{,i}\eta_{,j} + \rho_{,j}\eta_{,i} + \beta\rho_{,i}\rho_{,j}.$$

From this it follows that

$$(6.32) \qquad \zeta' = 0;$$

otherwise g_{ij} would be of rank $\leqq 2$. Consequently, (6.31) reduces to $\eta_{,[j}\rho_{,k]} = 0$, which shows that *η is a function of ρ alone.*

Moreover, on account of (6.32), we have from (6.30) that

$$(6.33) \qquad L = \zeta = \omega = 0.$$

Therefore, equations (6.12) and (6.1) become

$$(6.34) \qquad L_{ij} = \eta(\rho)\rho_{,i}\rho_{,j},$$

$$(6.35) \qquad \rho_{,ij} = -\phi\eta\rho_{,i}\rho_{,j},$$

where $\eta \neq 0$; otherwise C_n would be an S_n.

Equation (6.34) proves the necessity of condition (6.4) for the case $\xi = 0$. That this condition is also sufficient will now be proved by an argument similar to that for the case $\xi \neq 0$. Let there be given a C_n whose tensor L_{ij} is of the form (6.4). Using (6.4) in $(1.4)_2$, we deduce, as in (6.24),

$$(6.36) \qquad \bar{\rho}_{,ij} = \bar{\lambda}\bar{\rho}_{,i}\bar{\rho}_{,j}, \qquad \bar{\lambda} = \text{a scalar}.$$

The integrability condition of this is

$$(6.37) \qquad R^l_{i\,jk}\bar{\rho}_{,l} = \bar{\lambda}_{,[k}\bar{\rho}_{,j]}\bar{\rho}_{,i} + \bar{\lambda}\bar{\rho}_{,i[k}\bar{\rho}_{,j]}.$$

Now in consequence of $\xi = 0$, (6.4) and $(1.4)_1$, the left-hand member vanishes, while at the same time, the last term is zero because of (6.36). Hence (6.37) becomes $\bar{\lambda}_{,[k}\bar{\rho}_{,j]} = 0$, showing that $\bar{\lambda}$ *is a function of* $\bar{\rho}$ *alone*

Now by comparison of (6.34) and (6.4), it follows that $\bar{\rho} = \bar{\rho}(\rho)$. When this is substituted in (6.4) and (6.36), the result is

$$(6.38) \qquad L_{ij} = \bar{\eta}\bar{\rho}'^2\rho_{,i}\rho_{,j}, \qquad \bar{\rho}'\rho_{,ij} = (\bar{\lambda}\bar{\rho}'^2 - \bar{\rho}'')\rho_{,i}\rho_{,j}.$$

Comparing these with (6.34) and (6.35), we have

$$\eta = \bar{\eta}\bar{\rho}'^2, \qquad \bar{\rho}' = \frac{\bar{\lambda}\bar{\rho}'^2 - \bar{\rho}''}{-\phi\eta},$$

and hence

$$(6.39) \qquad \phi = \frac{\bar{\rho}'' - \bar{\lambda}\bar{\rho}'^2}{\bar{\eta}\bar{\rho}'^3},$$

which is identical with (6.28) and is therefore equivalent to (6.6). Equation (6.39), or (6.6), must be satisfied if the C_n in question permits a solution ρ of (6.1) such that $\xi = 0$. But on the other hand, if (6.39) is satisfied, equation (6.1), which is now $\rho_{,ij} + \phi L_{ij} = 0$ by (6.33), is a consequence of (6.38). Therefore we may conclude that for a C_n satisfying (6.4), the solution of (6.1) is identical with the solution $\rho = \rho(\bar{\rho})$ of equation (6.6). Hence Theorem 6.1 is completely established.

7. Applications

In §8 canonical forms for the fundamental tensors of the C_n considered in the preceding section will be obtained. But in the meantime, we apply the preceding results to a $C_n(a)$ which permits a solution of equation (3.1) or (4.6). For convenience we denote by $C_n(a)$ a C_n of constant curvature a. We first show that

THEOREM 7.1. *For a C_n whose tensor L_{ij} is of the form* (6.3) *or* (6.4), *equation* $(3.1)_1$ *admits infinitely many one-parameter family of solutions* $\rho(x^k, x^m)$ *such that* $\rho(x^k, x^m_o) = 1$ *for some value* x^m_o *of the parameter* x^m.

PROOF. By Theorem 6.1, in the case under consideration the solution of $(3.1)_1$ is identical with the solution $\rho = \rho(\bar{\rho})$ of

$$(7.1) \qquad \rho'' + \bar{\lambda}\rho' + \frac{\bar{\eta}}{n-2}(\rho^{n-1} - \rho) = 0,$$

where $\bar{\lambda}$ is a certain function of $\bar{\rho}$. According to the existence theorem, the solution of (7.1) can be written as

$$\rho = (\rho)_o + (\rho')_o(\bar{\rho} - \bar{\rho}_o) + \tfrac{1}{2}(\rho'')_o(\bar{\rho} - \bar{\rho}_o)^2 + \frac{1}{3!}(\rho''')_o(\bar{\rho} - \bar{\rho}_o)^3 + \cdots$$

where $\bar{\rho}_o$ is some value of $\bar{\rho}$; $(\rho)_o$ and $(\rho')_o$ are arbitrary constants; and $(\rho'')_o$, $(\rho''')_o$, etc. are determined from (7.1). If we take $(\rho)_o = 1$ and $(\rho')_o = 0$, then it is readily seen that $(\rho'')_o$, $(\rho''')_o$, etc. are all zero, and consequently the

corresponding solution is $\rho = 1$. Hence, if $f(x^m)$ and $h(x^m)$ are any two functions of x^m, the one-parameter family of solutions $\rho(x^k, x^m)$ of (7.1) corresponding to $(\rho)_o = f(x^m) - f(x_o^m) + 1$ and $(\rho')_o = h(x^m) - h(x_o^m)$ is such that $\rho(x^k, x_o^m) = 1$. Thus our theorem is proved.

As a direct consequence of Theorem 6.1, we have that[15]

THEOREM 7.2. *For a $C_n(a)$, not an S_n, equation* (4.6), *i.e.*

$$\rho_{,ij} + \rho L_{ij} = -\tfrac{1}{2}na\rho g_{ij}, \tag{7.2}$$

admits a solution satisfying $g^{ij}\rho_{,i}\rho_{,j} = 0$, if and only if the tensor L_{ij} of $C_n(a)$ is of the form (6.4).

Also by Theorem 6.1, if, for a $C_n(a)$, not an S_n, equation (7.2) admits a solution satisfying $g^{ij}\rho_{,i}\rho_{,j} \neq 0$, then the tensor L_{ij} of $C_n(a)$ must be of the form (6.3). This condition is also sufficient; indeed, we shall now prove

THEOREM 7.3. *For a $C_n(a)$, not an S_n, equation* (7.2) *admits a solution satisfying $g^{ij}\rho_{,i}\rho_{,j} \neq 0$, if and only if the tensor L_{ij} of $C_n(a)$ is of the form* (6.3). *If $C_n(a)$ satisfies this condition, the solution of* (7.2) *is*

$$\rho = A \exp.\left[-\int \frac{(\bar{\xi} + \frac{1}{2}na)\bar{\eta}'}{\bar{\xi}'}\, d\bar{\rho}\right], \qquad A = \text{const.}, \tag{7.3}$$

where the prime denotes differentiation with respect to $\bar{\rho}$.

To prove this theorem we need the formula

$$\bar{\chi}' - \bar{\lambda}\bar{\chi} + \bar{\xi} + \tfrac{1}{2}na = 0, \qquad \text{where} \qquad \bar{\chi} = \frac{\bar{\xi}'}{\bar{\eta}}, \tag{7.4}$$

and $\bar{\lambda} = \bar{\lambda}(\bar{\rho})$ is the function appearing in (6.5), namely,

$$\bar{\rho}_{,ij} = \bar{\lambda}\bar{\rho}_{,i}\bar{\rho}_{,j} + \bar{\chi}g_{ij}. \tag{7.5}$$

Equation (7.4) can be verified as follows. The integrability condition of (7.5) is

$$R^l_{i\,jk}\bar{\rho}_{,l} = \bar{\lambda}\bar{\rho}_{,i[k}\bar{\rho}_{,j]} + g_{i[j}\bar{\chi}_{,k]}\,.$$

In consequence of $(1.4)_1$, (6.3) and (7.6), the two members of the above equation reduce to

$$\frac{1}{n-2}(L_{i[j}\bar{\rho}_{,k]} + g_{i[j}L^l_{k]}\bar{\rho}_{,l}) = \frac{1}{n-2}(\bar{\xi}\bar{\eta} + 2\bar{\xi})g_{i[j}\bar{\rho}_{,k]}$$
$$= -(\bar{\xi} + \tfrac{1}{2}na)g_{i[j}\bar{\rho}_{,k]}\,,$$
$$\bar{\lambda}\bar{\chi}g_{i[k}\bar{\rho}_{,j]} + \bar{\chi}'g_{i[j}\bar{\rho}_{,k]} = (\bar{\chi}' - \bar{\lambda}\bar{\chi})g_{i[j}\bar{\rho}_{,k]}\,,$$

respectively. Equality of these two expressions gives (7.4).

We may now prove our theorem. From the argument below (6.29) it follows that for a $C_n(a)$ satisfying (6.3), equation (7.2) is equivalent to the two equa-

[15] Equation (7.2), for the case of a $C_3(O)$, has been discussed by Chou (**3**, **4**).

tions obtained by putting $\phi = \rho$, $\omega = -\frac{1}{2}na\rho$ in (6.29) and (6.6). These two equations are

$$-\rho' = \frac{(\bar{\zeta} + \frac{1}{2}na)\rho}{\bar{\chi}}, \tag{7.6}$$

$$\rho'' + \bar{\lambda}\rho' + \bar{\eta}\rho = 0. \tag{7.7}$$

Differentiating (7.6) and making use of (7.4) we have

$$\begin{aligned} -\rho'' &= \frac{(\bar{\zeta} + \frac{1}{2}na)\rho'}{\bar{\chi}} + \frac{\bar{\zeta}'\rho}{\bar{\chi}} - \frac{(\bar{\zeta} + \frac{1}{2}na)\rho\bar{\chi}'}{\bar{\chi}^2} \\ &= -\frac{(\bar{\zeta} + \frac{1}{2}na)^2\rho}{\bar{\chi}^2} + \bar{\eta}\rho - \frac{(\bar{\zeta} + \frac{1}{2}na)\rho\bar{\lambda}}{\bar{\chi}} + \frac{(\bar{\zeta} + \frac{1}{2}na)^2\rho}{\bar{\chi}^2} \\ &= \bar{\lambda}\rho' + \bar{\eta}\rho. \end{aligned}$$

Therefore, (7.7) is a consequence of (7.6), and hence the solution of (7.2) exists, being given by (7.6). Our theorem is thus proved.

8.1. Canonical form for the fundamental forms of some C_n

Let us now return to the general case and proceed to construct in certain privileged coordinates the fundamental forms of the C_n studied in §6. Consider first the case $\xi \neq 0$. If we put $\phi = 0$ in (6.17), we have $\omega = \frac{1}{2}\xi'$. Thus it follows from Theorem 6.1 for $\phi = 0$ that the tensor L_{ij} of a C_n is of the form (6.3), if and only if the equation

$$\rho_{,ij} = \tfrac{1}{2}\xi' g_{ij} \tag{8.1}$$

admits a solution ρ. Following Brinkmann (**2**, 123–4), we can prove that in any V_n for which equation (8.1) admits a solution ρ, a coordinate system exists in which $x^n = \rho$ and the fundamental form of V_n has the form

$$\xi\, h_{bc}(x^d)\, dx^b\, dx^c + \frac{(d\rho)^2}{\xi}, \qquad b, c, d = 1, \cdots, n-1, \tag{8.2}$$

where $h_{bc}(x^d)$ is independent of ρ. Conversely, for a V_n with fundamental form (8.2), where ξ is any function of ρ, ρ satisfies (8.1).

We first suppose that $\xi' \neq 0$. Put $z^{-2} = e\xi > 0$, where $e = \pm 1$, and let e be absorbed in $h_{bc}(x^d)$. Then the form (8.2) becomes

$$z^{-2}\left[h_{bc}(x^d)\, dx^b\, dx^c + \frac{(dz)^2}{Z}\right], \tag{8.3}$$

where we have written $Z = Z(z) = ez^{-4}\left(\frac{d\rho}{dz}\right)^{-2}$. Now *the condition for* (8.3) *to represent a* C_n *is that the form* $h_{bc}\, dx^b\, dx^c$ *represents an* S_{n-1}. In fact, (8.3) is conformal to the separable form

$$'g_{ij}\, dx^i\, dx^j = h_{bc}(x^d)\, dx^b\, dx^c + \frac{(dz)^2}{Z}, \tag{8.4}$$

and it was shown by Ficken (**9**, 897) that the condition for (8.4) to represent a C_n is equivalent to the condition for $h_{bc}\,dx^b\,dx^c$ to represent an S_{n-1}.

We now calculate the scalar curvature of (8.3). It is known that the scalar curvatures C and $'C$ of (8.3) and (8.4) are connected by (cf. 2.13))

$$C = 'Cz^2 + \frac{2}{n} z\,'g^{ij} z_{/ij} - 'g^{ij} z_{/i} z_{/j},$$

where the solidus denotes covariant differentiation with respect to (8.4). But $z_{/b} = 0$, $z_{/n} = 1$, and, if $'\Gamma^n_{nn}$ is a component of the Christoffel symbol of the second kind for (8.4),

$$'g^{ij} z_{/ij} = 'g^{nn} z_{/nn} = Z(-'\Gamma^n_{nn}) = \tfrac{1}{2}Z'.$$

Therefore we have

$$C = 'Cz^2 + \frac{1}{n} zZ' - Z. \tag{8.5}$$

To express $'C$ in terms of the scalar curvature B of h_{bc}, we note that the non-vanishing components of the Ricci tensor of (8.4) are $'R_{bc} = \overset{(h)}{R}_{bc}$, where the superindex (h) indicates that the tensor is refered to h_{bc}. Therefore it follows that $n'C = (n-2)B$. Using this in (8.5) we have

$$nC = (n-2)Bz^2 + zZ' - nZ = (n-2)Bz^2 + z^{n+1}(z^{-n}Z)'. \tag{8.6}$$

If h_{bc} represents an S_{n-1}, $B = \text{const.}$ and (8.6) shows that $C = C(z)$. Then, by integrating (8.6) we have

$$Z = Az^n + Bz^2 + nz^n \int Cz^{-(n+1)}\,dz, \tag{8.7}$$

where A is a constant.

We now suppose that $\xi' = 0$. Then since by hypothesis $\xi = \text{const.} \neq 0$, an obvious transformation would reduce the fundamental form (8.2) to the form

$$h_{bc}(x^d)\,dx^b\,dx^c + e\,(dz)^2, \qquad e = \pm 1.$$

Like (8.3), this represents a $C_n(C)$ if and only if h_{bc} represents an $S_{n-1}(B)$; and then $nC = (n-2)B = \text{const.}$

Summing up the preceding results and using the Riemann form for the fundamental form of an S_{n-1}, we have

THEOREM 8.1. *A V_n is a C_n whose tensor L_{ij} is of the form* (6.3) *if and only if its fundamental form can be reduced to either of the following forms*:

$$\frac{\sum_a e_a(dx^a)^2}{z^2\left[1 + \frac{B}{4}\sum_a e_a(x^a)^2\right]^2} + \frac{(dz)^2}{z^2\left[Az^n + Bz^2 + nz^n \int Cz^{-(n+1)}\,dz\right]}, \tag{8.8}$$

$$\frac{\sum_a e_a(dx^a)^2}{\left[1 + \frac{B}{4} \sum_a e_a(x^a)^2\right]^2} + e_n(dz)^2, \tag{8.9}$$

where $a = 1, \cdots, n-1$; *each* e *is* ± 1; A *and* B *are constants; and* $C = C(z)$. *The scalar curvature of* C_n *is* C *for* (8.8) *and is* $\frac{n-2}{n} B$ *for* (8.9).

In particular, when $C = \text{const.}$, we have

COROLLARY 8.1. *A* V_n *is a* C_n *whose tensor* L_{ij} *is of the form* (6.3) *and whose scalar curvature* C *is constant, if and only if its fundamental form can be reduced to either of the following forms*:

$$\frac{\sum_a e_a(dx^a)^2}{z^2\left[1 + \frac{B}{4} \sum_a e_a(x^a)^2\right]^2} + \frac{(dz)^2}{z^2(Az^n + Bz^2 - C)} \tag{8.8'}$$

$$\frac{\sum_a e_a(dx^a)^2}{\left[1 + \frac{nC}{4(n-2)} \sum_a e_a(x^a)^2\right]^2} + e_n(dz)^2. \tag{8.9'}$$

We remark that *in* (8.8) *or* (8.8′), *if the constant* A *is not zero, it may be supposed to be* ± 1. In fact, if we put $z = \alpha' z$, $\alpha = \text{const.}$, (8.3) and (8.7) become

$$'z^{-2}\left[\alpha^{-2} h_{bc}(x^d)\, dx^b\, dx^c + \frac{(d'z)^2}{'Z}\right],$$

$$'Z = A\alpha^n\, 'z^n + B\alpha^2\, 'z^2 + n\, 'z^n \int C\, 'z^{-(n+1)}\, d'z.$$

Now if $A \neq 0$, α can be so chosen that $A\alpha^n = \pm 1$. But since the scalar curvature of $(\alpha^{-2}h_{bc})$ is equal to α^2 times that of h_{bc}, this transformation from z to $'z$ does not affect the property of h_{bc} being the fundamental tensor of an $S_{n-1}(B)$. Hence our assertion is proved.

8.2. Continuation

We now consider the case $\xi = 0$. By Theorem 6.1, the tensor L_{ij} of a $C_n(0)$ is of the form (6.4) if and only if the equations

$$\rho_{,ij} = 0, \qquad \xi = g^{ij}\rho_{,i}\rho_{,j} = 0 \tag{8.10}$$

admit a solution for ρ. Then ρ is a function of $\bar{\rho}$ and we have

$$L_{ij} = \eta(\rho)\rho_{,i}\rho_{,j}, \qquad \eta \neq 0. \tag{8.11}$$

Brinkmann (**2,** 131–2) proved that in consequence of (8.10) a coordinate system x^p, y, z ($p, q, r, s = 1, \cdots, n-2$) exists in which $\rho = z$ and the fundamental form of $C_n(0)$ has the form

$$g_{ij}\, dx^i\, dx^j = f_{pq}\, dx^p\, dx^q + 2m_p\, dx^p dz + m(dz)^2 + 2\, dy\, dz, \tag{8.12}$$

where the functions f_{pq}, m_p and m are independent of y. For (8.12), the non-vanishing components of Γ^k_{ij} are

$$\Gamma^r_{pq} = \overset{(f)}{\Gamma^r_{pq}}, \quad \Gamma^r_{pn}, \quad \Gamma^r_{nn}, \quad \Gamma^{n-1}_{pq}, \quad \Gamma^{n-1}_{nn}, \quad \Gamma^{n-1}_{pn-1}, \quad \Gamma^{n-1}_{pn},$$

where the superindex (f) indicates that the quantities are referred to the form $f_{pq}\, dx^p\, dx^q$, z being considered as a parameter. From these it follows that

$$R^s_{pqr} = \partial_{[q}\Gamma^s_{r]p} - \Gamma^i_{p[q}\Gamma^s_{r]i} = \overset{(f)}{R^s_{pqr}}. \tag{8.13}$$

On the other hand, since $C_n(0)$ is conformal-Euclidean, we have from $(1.4)_1$ and (8.11) that

$$R^l_{i\,jk} = Z(z_{,i}z_{,[j}\delta^l_{k]} + g_{i[j}z_{,k]}z_{,}{}^l), \tag{8.14}$$

where $z_{,}{}^l = g^{li}z_{,i}$, and Z is defined by $\eta(\rho) \equiv (n-2)Z(z)$. But

$$z_{,p} = z_{,n-1} = 0, \qquad z_{,n} = 1; \qquad z_{,}{}^p = z_{,}{}^n = 0, \qquad z_{,}{}^{n-1} = 1. \tag{8.15}$$

Therefore the components R^s_{pqr} of $R^l_{i\,jk}$ are zero. Thus (8.13) gives $\overset{(f)}{R^s_{pqr}} = 0$, and consequently, for each value of z the form $f_{pq}\, dx^p\, dx^q$ in (8.12) represents a Euclidean V_{n-2}. Hence by means of a suitable transformation of the form

$$'x^p = {}'x^p(x^r, z), \qquad 'y = y, \qquad 'z = z,$$

the fundamental form (8.12) can be reduced to (after dropping the dash)

$$g_{ij}\, dx^i\, dx^j = \sum_p e_p(dx^p)^2 + 2m_p\, dx^p\, dz + m(dz)^2 + 2\, dy\, dz, \quad e^p = \pm 1, \tag{8.16}$$

where, as in (8.12), the functions m_p and m are independent of y.

For this fundamental form we have

$$\Gamma^s_{pq} = \Gamma^s_{qn-1} = \Gamma^s_{nn-1} = \Gamma^n_{pq} = \Gamma^n_{pn} = 0, \qquad \Gamma^s_{pn} = \tfrac{1}{2}e_s\partial_{[p}m_{s]},$$

and whence $R^s_{pqn} = \partial_{[q}\Gamma^s_{n]p} - \Gamma^i_{p[q}\Gamma^s_{n]i} = \partial_q\Gamma^s_{np} = \frac{1}{2}e_s\partial_q(\partial_{[p}m_{s]})$. But R^s_{pqn} should be zero by (8.14) and (8.15). Therefore we have

$$\partial_{[p}m_{s]} = P_{ps}(z) = -P_{sp}(z), \tag{8.17}$$

where $P_{ps}(z)$ are some functions of z alone.

It is easily seen that if $Y(x^r, z)$ is a function of x^r and z, and $Q^p_r(z)$ are functions of z satisfying the equations

$$\sum_r e_r Q^p_r Q^q_r = e_p\, \delta_{pq}, \tag{8.18}$$

or their equivalent

$$\sum_p e_p Q^p_r Q^p_q = e_r\, \delta_{rq}, \tag{8.18'}$$

then the following change of coordinates

$$'x^p = Q^p_r(z)x^r, \qquad 'y = y + Y(x^r, z), \qquad 'z = z \tag{8.19}$$

will preserve the form of (8.16). Now for (8.16) the contravariant components g^{ij} of g_{ij} are

$$
\begin{aligned}
&g^{np} = 0, \qquad g^{nn-1} = 1, \qquad g^{nn} = 0; \\
&g^{n-1n-1} = g^{n-1n-1}, \qquad g^{pn-1} = -e_p m_p, \qquad g^{pq} = e_p \delta_{pq}.
\end{aligned} \tag{8.20}
$$

Therefore if we denote by $'g^{pn-1}$ and $'m_p$ the corresponding values to g^{pn-1} and m_p in this new coordinate system (8.19), we have

$$
\begin{aligned}
-e_p\, 'm_p = {}'g^{pn-1} &= g^{ij} \frac{\partial}{\partial x^i}(Q^p_r x^r) \frac{\partial}{\partial x^j}(y + Y) \\
&= \sum_r e_r Q^p_r \frac{\partial Y}{\partial x^r} - \sum_r e_r m_r Q^p_r + (Q^p_r)' x^r.
\end{aligned} \tag{8.21}
$$

We shall now prove that the functions Y and Q^p_r may be so chosen that $'m_p = 0$.

Making use of the conjugate system $\bar{Q}^q_p$ of Q^p_r defined by

$$Q^p_r \bar{Q}^q_p = \delta^q_r, \tag{8.22}$$

it follows from (8.21) that the condition $'m_p = 0$ is equivalent to

$$\frac{\partial Y}{\partial x^q} - m_q + e_q \bar{Q}^q_p (Q^p_r)' x^r = 0, \qquad (q \text{ not summed}).$$

But we have from (8.18) and (8.22) that $e_p \bar{Q}^q_p = e_q Q^p_q$, therefore the above equations can be written

$$\frac{\partial Y}{\partial x^q} = m_q - \sum_p e_p Q^p_q (Q^p_r)' x^r. \tag{8.23}$$

Now let us consider the variable z in Y as a parameter. Then the integrability condition of (8.23) is

$$\partial_{[r} m_{q]} = \sum_p e_p Q^p_{[q} (Q^p_{r]})',$$

which becomes, by (8.17),

$$\sum_p e_p Q^p_{[q} (Q^p_{r]})' = P_{rq}. \tag{8.24}$$

If we write $\bar{n} = n - 2$, then (8.24) are $\frac{1}{2}\bar{n}(\bar{n} - 1)$ equations, which, together with the $\frac{1}{2}\bar{n}(\bar{n} + 1)$ equations (8.18), form a system of $\bar{n}^2$ equations for the $\bar{n}^2$ unknowns Q^p_r. That this system of equations is compatible we shall prove later. For the moment, we assume that a set of values Q^p_r have been found satisfying (8.18) and (8.24). Then, equations (8.23) can be solved for Y. Hence the functions Q^p_r and Y in (8.19) can be so determined that $'m_p = 0$, as was to be proved. In other words, there exists a coordinate system in which the fundamental form of $C_n(0)$ has the form

$$g_{ij}\, dx^i\, dx^j = \sum_p e_p (dx^p)^2 + m(dz)^2 + 2\, dy\, dz. \tag{8.25}$$

Now we have to determine m so that (8.14) is satisfied. For (8.25) the non-vanishing components of Γ^k_{ij} and R^l_{ijk} are[16]

$$\Gamma^p_{nn} = -\tfrac{1}{2}e_p\partial_p m, \qquad \Gamma^{n-1}_{nn} = -\tfrac{1}{2}\partial_n m, \qquad \Gamma^{n-1}_{pn} = \tfrac{1}{2}\partial_p m;$$

$$R^s_{nnr} = \tfrac{1}{2}e_s\partial_r\partial_s m, \qquad R^{n-1}_{pqn} = \tfrac{1}{2}\partial_q\partial_p m.$$

On the other hand, we have from (8.14) and (8.15) that the only non-vanishing components of R^l_{ijk} are

$$R^s_{nnr} = Z\delta^s_r\,, \qquad R^{n-1}_{pqn} = Zg_{pq}\,.$$

Comparison of these two different expressions for the components of R^l_{ijk} shows that

$$\partial_q\partial_p m = 2Zg_{pq} = 2Ze_p\delta_{pq}\,,$$

which gives, since m is independent of y,

$$m = Z\sum_p e_p(x^p)^2 + Z_p x^p + Z_{n-1}\,, \tag{8.26}$$

where the Z's are some functions of z alone.

We can prove, by direct computations, that the fundamental form (8.25) with the value (8.26) for m is actually the fundamental form of a $C_n(0)$ whose tensor L_{ij} is of the form (6.4). Indeed, for this fundamental form the only non-vanishing components of R^l_{ijk} and R_{ij} are

$$R^s_{nnr} = Z\delta^s_r\,, \qquad R^{n-1}_{pqn} = Zg_{pq}\,; \qquad R_{nn} = (n-2)Z.$$

From these it follows that $L = R = 0$, $L_{ij} = R_{ij}$, and consequently, we can easily verify that conditions (1.4) are satisfied.

Hence we may state the following

THEOREM 8.2. *A V_n is a $C_n(0)$ whose tensor L_{ij} is of the form (6.4) if and only if its fundamental form can be reduced to the form*

$$\sum_p e_p(dx^p)^2 + 2\,dy\,dz + \Big[Z\sum_p e_p(x^p)^2 + Z_p x^p + Z_{n-1}\Big](dz)^2, \tag{8.27}$$

$$e_p = \pm 1, \qquad p = 1, \cdots, n-2,$$

where the Z's are some functions of z alone.

To complete the proof of this theorem, we now show, as was asserted, that equations (8.18) and (8.24) can actually be solved for Q^p_r. From the equivalent (8.18′) of (8.18) we have

$$\Big(\sum_p e_p Q^p_r Q^p_q\Big)' = \sum_p e_p(Q^p_r)'(Q^p_q) + \sum_p e_p Q^p_r(Q^p_q)' = 0. \tag{8.28}$$

Hence (8.24) can be written

$$2\sum_p e_p(Q^p_q)(Q^p_r)' = P_{rq}\,.$$

Transvecting these by $\bar{Q}^q_s$, we find

$$2e_s(Q^s_r)' = P_{rq}\bar{Q}^q_s \qquad (s \text{ not summed}),$$

that is, $\quad (Q^s_r)' = \tfrac{1}{2}e_s\bar{Q}^q_s P_{rq} = \tfrac{1}{2}e_s\sum_q e_p e_s Q^s_q P_{rq} \quad (s \text{ not summed}).$

---

[16] Strictly speaking, there are two more non-vanishing components of R^l_{ijk}, namely, R^s_{nrn} $(= -R^s_{nnr})$ and R^{n-1}_{pnq} $(= -R^{n-1}_{pqn})$.

Therefore, on account of (8.18), equations (8.24) are equivalent to

$$(8.29) \qquad (Q_r^s)' = \tfrac{1}{2} \sum_q e_q Q_q^s P_{rq} .$$

These are a system of $\bar{n}^2 = (n - 2)^2$ ordinary linear differential equations in the $\bar{n}^2$ unknowns Q_r^s and one independent variable z. They have a unique solution for any set of initial values $(Q_r^s)_0$ at $z = z_0$. We can easily prove that when $(Q_r^s)_0$ are suitably chosen, the solution Q_r^s satisfies (8.18). Indeed, since $(Q_r^s)_0$ can always be found satisfying (8.18), we need only show that the expression (8.28) is satisfied because of (8.29). Now

$$\begin{aligned}(\sum_p e_p Q_r^p Q_q^p)' &= \sum_p e_p(Q_q^p)(\tfrac{1}{2}\sum_s e_s Q_s^p P_{rs}) + \sum_p e_p(Q_r^p)(\tfrac{1}{2}\sum_s e_s Q_s^p P_{qs}) \\ &= \tfrac{1}{2}(\sum_s e_q \delta_{qs} e_s P_{rs} + \sum_s e_r \delta_{rs} e_s P_{qs}) \\ &= \tfrac{1}{2}(P_{rq} + P_{qr}),\end{aligned}$$

which is zero by (8.17). Hence our assertion is proved.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

REFERENCES

1. BRINKMANN, H. W., "*Riemann spaces conformal to Einstein spaces,*" Math. Ann., 91 (1924), 269–278.
2. BRINKMANN, H. W., "*Einstein spaces which are mapped conformally on each other,*" Math. Ann., 94 (1925), 119–145.
3. CHOU, P. Y., "*Isotropic static solutions of the field equation in Einstein's theory of gravitation,*" American J. Math., 59 (1937), 754–763.
4. CHOU, P. Y., "*On the method of finding isotropic static solutions of Einstein's field equations of gravitation,*" American J. Math., 62 (1940), 43–48.
5. EISENHART, L. P., Riemannian geometry (1926), Princeton.
6. EISENHART, L. P., Continuous groups of transformations (1934), Princeton.
7. FIALKOW, A., "*Totally geodesic Einstein spaces,*" Bull. American Math. Soc., 45 (1939), 423–428.

7′. FIALKOW, A., "*Correction to 'Totally geodesic Einstein spaces,'*" loc. cit., 48 (1942), 167–168.

8. FIALKOW, A., "*Conformal geodesics,*" Trans. American Math. Soc., 45 (1939), 443–473.
9. FICKEN, F. A., "*The Riemannian and affine differential geometry of product-spaces,*" Ann. Math., 40 (1939), 892–913.
10. SCHOUTEN, J. A., AND STRUIK, D. J., "*On some sroperties of general manifolds relating to Einstein's theory of graviation,*" American J. Math., 43 (1921), 217–221.
11. SCHOUTEN, J. A., AND STRUIK, D. J., Einführung in die neueren Methoden der Differentialgeometrie, II, (1938), Groningen.
12. YANO, K., "*Concircular geometry, I. Concircular transformations,*" Proc. Imp. Acad. Tokyo, 16 (1940), 195–200.
13. YANO, K., "*Concircular geometry, II. Integrability conditions of $\rho_{\mu\nu} = \phi g_{\mu\nu}$,*" loc. cit., 354–360.
14. WONG, YUNG-CHOW, "*Some Einstein spaces with conformally separable fundamental tensors,*" in course of publication in the Transactions of the American Mathematical Society.

ANNALS OF MATHEMATICS
Vol. 44, No. 2, April, 1943

EUCLIDEAN APPLICATIONS OF THE PROJECTIVE DIFFERENTIAL GEOMETRY OF THE R_λ-CORRESPONDENT

BY P. O. BELL AND W. C. FOREMAN

(Received July 3, 1942)

INTRODUCTION

Let C_λ denote an arbitrary curve on a surface S in ordinary projective space. P. O. Bell has recently presented the following construction for the line which he calls the *R_λ-correspondent*[1] of the tangent to C_λ at a point x: Let X denote a point of C_λ distinct from x, let U^1, U^2 denote, respectively, the points of intersection of the asymptotic u^1- and u^2-curves passing through x with the asymptotic u^2- and u^1-curves passing through X, and let W denote the point of intersection of the tangent plane to S at x with the line joining the points U^1, U^2. *The R_λ-correspondent of the tangent to C_λ at x is the limit of the line joining x and W as X approaches x along C_λ.* He has proved, moreover, the following three theorems:

(i) A curve C_λ is a curve of Darboux if and only if at each of its points the R_λ-correspondent of the tangent to C_λ coincides with this tangent;

(ii) A curve C_λ is a curve of Segre if and only if at each of its points the tangent to C_λ and its R_λ-correspondent are conjugate tangents of S;

(iii) A curve C_λ is a curve of Segre if and only if for a general point x of C_λ the limit of W as X tends to x along C_λ is a point W_0 distinct from x. The point W_0 is the intersection of the directrix of the first kind of Wilczynski with the tangent at x to the corresponding curve C_λ of Darboux.

The correspondent whose definition results from that of the R_λ-correspondent upon the removal of the restriction that the asymptotic net be parametric will hereafter be called the R_λ-correspondent. This paper is concerned largely with the study of this correspondent and of a dual correspondent which will be called the R^*_λ-correspondent. In chapter I the study is from the point of view of projective differential geometry. The determination of the R_λ- correspondent for a general parametric net leads naturally to projective definitions of a number of interesting nets as well as of a rectilinear congruence covariantly determined with respect to the parametric net. Among the nets thus determined are the R_λ-net, the principal-associate net of the parametric net and the associate conjugate net of the parametric net. In Chapter II euclidean geometric characterizations of the R_λ- and R^*_λ-correspondents are first obtained. These characterizations form the basis for the formulation of dual theories in euclidean differential geometry of nets. A unique parametric net is discovered which is such that every associated R_λ-net is orthogonal. This net consists of curves

[1] P. O. Bell, Bull. Am. Math. Soc. 47, 1941, pp. 509–511. For a detailed study of the R_λ-correspondent and of associated R_λ-derived curves for the case in which the asymptotic curves are parametric see Trans. Am. Math. Soc. 46.

whose normal curvatures at a point are equal to the mean of the principal normal curvatures at the point. We call these curves the lines of mean curvature of S. The dual of this net is the unique conjugate parametric net which is such that every associated R_λ^*-net is conjugate. This net is the mean conjugate net and consists of those curves of S whose radii of normal curvature at a point x are equal to the mean of the principal radii of normal curvature at the point. Finally euclidean characterizations are obtained for the curves of Segre and the curves of Darboux and simple new determinations are given for the lines of curvature.

I. Projective Differential Geometry

1. The R_λ-correspondent

Let the homogeneous coordinates of a general point x of an analytic surface S and of a corresponding point z not in the tangent plane to S at x be functions x^i and z^i $(i = 1, 2, 3, 4)$, of two independent variables u^1, u^2. The pairs of functions (x, z) are solutions of a system of differential equations of the form[2]

$$\begin{aligned} \frac{\partial^2 x}{\partial u^i \partial u^j} &= \left\{ {k \atop ij} \right\} \frac{\partial x}{\partial u^k} + a_{ij} x + d_{ij} z, \\ \frac{\partial z}{\partial u^i} &= b_i^k \frac{\partial x}{\partial u^k} + c_i x + e_i z, \qquad (i, j, k = 1, 2), \end{aligned} \tag{1.1}$$

whose coefficients are functions of u^1, u^2 which satisfy certain integrability conditions. The usual summation convention of tensor analysis is employed throughout this paper.

The asymptotic curves are the integral curves of the differential equation

$$d_{ij} du^i du^j = 0. \tag{1.2}$$

An arbitrary one parameter family F_λ of curves is defined by the curvilinear equation

$$du^2 - \lambda du^1 = 0, \tag{1.3}$$

where λ is an arbitrary function of u^1, u^2. Let C_λ denote the curve of the family which passes through the point x. The curvilinear coordinates of a general point on a curve of S may be represented by the functions $u^i = u^i(t)$, $(i = 1, 2)$, of a parameter t. If this curve passes through x it will be the curve C_λ if for a general value of t

$$\lambda = \frac{du^2}{dt} \Big/ \frac{du^1}{dt}. \tag{1.4}$$

Let us determine the direction of the R_λ-correspondent as defined in the Introduction for a general non-asymptotic parametric net. The curvilinear coordinates of the point X are given by $u^1(t + \Delta t)$, $u^2(t + \Delta t)$. The points

[2] Cf. E. P. Lane, *Projective Differential Geometry of Curves and Surfaces* (hereafter referred to as P. D. G.), University of Chicago Press, 1932, p. 183, equations (35).

U^1 and U^2 are therefore given by $(u^1(t + \Delta t), u^2)$ and $(u^1, u^2(t + \Delta t))$. The general homogeneous coordinates of the points U^i are consequently functions of t, and may be represented by the development

$$
\begin{aligned}
U^i = x + \frac{\partial x}{\partial u^i}\frac{du^i}{dt}\Delta t &+ \left(\frac{\partial^2 x}{\partial u^{i2}}\left(\frac{du^i}{dt}\right)^2 + \frac{\partial x}{\partial u^i}\frac{d^2u^i}{dt^2}\right)(\Delta t)^2/2 \\
(1.5)\qquad &+ \left(\frac{\partial^3 x}{\partial u^{i3}}\left(\frac{du^i}{dt}\right)^3 + 3\frac{\partial^2 x}{\partial u^{i2}}\frac{du^i}{dt}\frac{d^2u^i}{dt^2} + \frac{\partial x}{\partial u^i}\frac{d^3u^i}{dt^3}\right)(\Delta t)^3/6 \\
&+ \left(\frac{\partial^4 x}{\partial u^{i4}}\left(\frac{du^i}{dt}\right)^4 + 6\frac{\partial^3 x}{\partial u^{i3}}\left(\frac{du^i}{dt}\right)^2\frac{d^2u^i}{dt^2} + f_1^i\frac{\partial^2 x}{\partial u^{i2}} + f_2^i\frac{\partial x}{\partial u^i}\right)(\Delta t)^4/24 + \cdots,
\end{aligned}
$$

wherein f_j^i, $(i, j = 1, 2)$, represent functions of u^1, u^2 which for our purposes do not require explicit determination, and where i is not summed.

By differentiating equations (1.1) we find that the coefficient of z in the expression for $\dfrac{\partial^3 x}{\partial u^{i3}}$ is $\dfrac{\partial d_{ii}}{\partial u^i} + d_{ii}e_i + d_{ij}\left(\begin{smallmatrix} j \\ ii\end{smallmatrix}\right)$. The coefficient of z in the expression for the homogeneous coordinates of the point U^i is, therefore $d_{ii}\left(\dfrac{du^i}{dt}\right)^2(\Delta t)^2/2 + \omega_i\left(\dfrac{du^i}{dt}\Delta t\right)^3/6 + \cdots$, wherein $\omega_i = \dfrac{\partial d_{ii}}{\partial u^i} + d_{ii}e_i + d_{ij}\left(\begin{smallmatrix} j \\ ii\end{smallmatrix}\right) + 3d_{ii}\left(\dfrac{d^2u^i}{dt^2}\right)\Big/\left(\dfrac{du^i}{dt}\right)^2$.
The point W has homogeneous coordinates which may be obtained by forming a linear combination of those of U^1 and U^2 which contains no z term. Such a combination is

$$
\left(3d_{22}\left(\frac{du^2}{dt}\right)^2 + \omega_2\left(\frac{du^2}{dt}\right)^3\Delta t\right)U^1 - \left(3d_{11}\left(\frac{du^1}{dt}\right)^2 + \omega_1\left(\frac{du^1}{dt}\right)^3\Delta t\right)U^2.
$$

Expanding this, we obtain the expression

$$
\begin{aligned}
&3\left(d_{22}\left(\frac{du^2}{dt}\right)^2 - d_{11}\left(\frac{du^1}{dt}\right)^2\right)x + \left(\omega_2\left(\frac{du^2}{dt}\right)^3 - \omega_1\left(\frac{du^1}{dt}\right)^3\right)(\Delta t)x \\
(1.6)\qquad &+ 3(\Delta t)\left(d_{22}\left(\frac{du^2}{dt}\right)^2\left(\frac{du^1}{dt}\right)\frac{\partial x}{\partial u^1} - d_{11}\left(\frac{du^1}{dt}\right)^2\left(\frac{du^2}{dt}\right)\frac{\partial x}{\partial u^2}\right) \\
&+ \text{terms of order } (\Delta t)^2
\end{aligned}
$$

for the homogeneous coordinates of the point W. Therefore, if the parametric curves are not asymptotic curves, the direction of the R_λ-correspondent of the tangent to C_λ at x is defined by

$$
du^2/du^1 = -d_{11}/d_{22}\lambda. \tag{1.7}
$$

Let C_W denote the curve described by the point W as Δt varies. The expression (1.6) for the coordinates of W shows that *if the parametric net is not the asymptotic net and the curve C_λ is not an integral curve of the equation*

$$
d_{11}(du^1)^2 - d_{22}(du^2)^2 = 0 \tag{1.8}
$$

[3] P. O. Bell, loc. cit.

then the R_λ-correspondent of the tangent to C_λ at x is tangent at x to the corresponding curve C_W, that is to say, the point x is the limit of W as Δt tends to zero.

If C_λ is an integral curve of (1.8), the limit of W as Δt tends to zero is a point W_0, distinct from x, whose homogeneous coordinates are given by

$$(1.9)\quad \left(\omega_2\left(\frac{du^2}{dt}\right)^3 - \omega_1\left(\frac{du^1}{dt}\right)^3\right)x + 3\left(d_{22}\left(\frac{du^2}{dt}\right)^2\frac{du^1}{dt}\frac{\partial x}{\partial u^1} - d_{11}\left(\frac{du^1}{dt}\right)^2\frac{du^2}{dt}\frac{\partial x}{\partial u^2}\right),$$

wherein $d_{22}\left(\frac{du^2}{dt}\right)^2 = d_{11}\left(\frac{du^1}{dt}\right)^2$. If we evaluate (1.9) explicitly for the directions $\lambda = \pm(d_{11}/d_{22})^{\frac{1}{2}}$, we obtain, except for unessential factors

$$(1.10)\quad \begin{aligned} &3\frac{\partial x}{\partial u^1} - \left(\frac{\partial \log (d_{11}d_{22}^3)^{\frac{1}{4}}}{\partial u^1} + e_1 + \frac{d_{1j}({}_1{}^j{}_1)}{d_{11}}\right)x \\ &\pm\left(\frac{d_{11}}{d_{22}}\right)^{\frac{1}{2}}\left(3\frac{\partial x}{\partial u^2} - \left(\frac{\partial \log (d_{11}^3 d_{22})^{\frac{1}{4}}}{\partial u^2} + c_2 + \frac{d_{2j}({}_2{}^j{}_2)}{d_{22}}\right)x\right). \end{aligned}$$

These two points W_0 determine a line which will be called the w-line associated with the parametric net at x.

For the parametric asymptotic net since $d_{11} = d_{22} = 0$ and $d_{12} \neq 0$, the directions of the R_λ-correspondent of the tangent to C_λ at x is determined by the terms of order $(\Delta t)^2$ in (1.6). This direction, which is defined by

$$(1.11)\qquad du^2/du^1 = -({}_1{}^2{}_1)/({}_2{}^1{}_2)\lambda^2$$

has been studied on previous occasions[3] from a projective standpoint.

For an arbitrarily selected parametric net there exists in correspondence with any one parameter family of curves F_λ of S a family of curves $F_{\bar\lambda}$ characterized by the property that at a general point x of S the tangent to the curve $C_{\bar\lambda}$ through x is the R_λ-correspondent of the tangent to C_λ at x. The family $F_{\bar\lambda}$, thus defined in association with a family F_λ and a given parametric net, will be called the *family of R_λ-derived curves.* If the parametric net is not the asymptotic net, equation (1.7), with u^1, u^2 allowed to vary, defines the family of R_λ-derived curves. If, however, the asymptotic net is parametric, equation (1.11) defines the family of R_λ-derived curves. The net which consists of the family F_λ and the family of R_λ-derived curves will be called the R_λ-net.

2. Associate nets of a parametric net

The correspondence defined by (1.7), between the direction of the tangent to C_λ at x and its R_λ-correspondent is an involution. At a point x of S, the double elements of this involution are the solutions of the equation

$$(2.1)\qquad d_{ii}(du^i)^2 = 0.$$

The net of integral curves of this equation will be called the *principal associate net* of the parametric net. Since the directions at x of the curves of this net are double elements of the involution (1.7), we have immediately

THEOREM (2.1) *At a general point x of S the tangents of the principal associate net of the parametric net are harmonic conjugates with respect to the tangent at x to an arbitrary curve C_λ through x and the R_λ-correspondent of this tangent.*

If $d_{12} = 0$, the integral curves of (2.1) are the asymptotic curves of S. Hence we have

THEOREM (2.2) *The asymptotic net is the principal associate net of an arbitrary conjugate parametric net.*

The unique conjugate net whose tangents at a general point x separate harmonically the tangents at x to the parametric curves is called the *associate conjugate net* of the parametric net. The differential equation of this net is equation (1.8). Hence we have

THEOREM (2.3) *The associate conjugate net of a parametric net consists of those curves, a general one C_λ of which has the property that, the limit of the point W as X tends toward x along C_λ is a point W_0 distinct from x. The two points W_0 which correspond to the two curves of the net which pass through x determine the w-line of the parametric net at x.*

The necessary and sufficient condition that the tangent to C_λ at x and its R_λ-correspondent be conjugate tangents is that the harmonic invariant of the two forms

$$d_{ij}du^i du^j \quad \text{and} \quad d_{11}\lambda(du^1)^2 + (d_{22}\lambda^2 - d_{11})du^1du^2 - d_{22}(du^2)^2$$

vanish; that is

$$d_{12}(d_{22}\lambda^2 - d_{11}) = 0. \tag{2.2}$$

If $d_{12} = 0$, equation (2.2) is satisfied independently of λ. Hence we have that

THEOREM (2.4) *Every R_λ-net derived in association with an arbitrary conjugate parametric net is a conjugate net.*

The equation $d_{22}\lambda^2 - d_{11} = 0$ is not only the equation for the directions at x of the curves of the associate conjugate net of the parametric net, but it is the necessary and sufficient condition that the tangents at x of the R_λ-net separate harmonically the tangents to the parametric net. Hence we have

THEOREM (2.5) *Corresponding to a non-asymptotic parametric net there is a unique R_λ-net whose tangents at a general point x separate harmonically the tangents to the parametric curves at the point. This R_λ-net is the associate conjugate net of the parametric net.*

If C_λ is an asymptotic curve of S it is clear from equation (1.2) that the direction of the other asymptotic curve at a general point x of C_λ is given by $\dfrac{du^2}{du^1} = \dfrac{d_{11}}{d_{22}\lambda}$. We have, therefore

THEOREM (2.6) *The tangents at x to the parametric curves of S separate harmonically the R_λ-correspondent of the tangent at x to an asymptotic curve C_λ and the tangent to the other asymptotic curve at x.*

The harmonic conjugate of the tangent at x to an arbitrary curve C_λ of S with respect to the tangents at x to the curves of any selected parametric net

may be geometrically characterized in the following simple manner. Let x and X denote two points on a curve C_λ of S, and let U^1 and U^2 denote the points of intersection of the parametric u^1- and u^2-curves passing through x with the parametric u^2- and u^1-curves passing through X. *The limit of the line joining U^1 and U^2 as X approaches x along C_λ is the harmonic conjugate of the tangent to C_λ at x with respect to the tangents at x to the parametric curves of S.*

The validity of the above construction may be proved as follows. The curvilinear coordinates of the points U^1 and U^2 are $\bar{u}^1$, u^2 and u^1, $\bar{u}^2$, respectively, wherein $\bar{u}^1(t) = u^1(t + \Delta t)$, $\bar{u}^2(t) = u^2(t + \Delta t)$. The coordinates u^1, $\bar{u}^2$ of U^2 may be defined in terms of those of U^1 by the relations $u^1(t) = \bar{u}^1(t - \Delta t)$, $\bar{u}^2(t) = u^2(t + \Delta t)$. Except for terms of order $(\Delta t)^2$ these coordinates are $\bar{u}^1 - \frac{du^1}{dt} \Delta t$, $u^2 + \frac{du^2}{dt} \Delta t$. Hence, the direction of the limit of the line joining U^1, U^2 as Δt tends to zero is given by $\frac{du^2}{du^1} = -\frac{du^2}{dt} \Big/ \frac{du^1}{dt} = -\lambda$. In view of the cross ratio equation $(0, \infty, \lambda, -\lambda) = -1$ the proof is complete.

As an important special case of the above construction we have

THEOREM (2.7) *If the asymptotic net is parametric, the limit of the line joining U^1, U^2 as X tends to x along C_λ is the conjugate of the tangent to C_λ at x.*

II. APPLICATIONS TO EUCLIDEAN DIFFERENTIAL GEOMETRY

3. Euclidean duality

In this section the groundwork will be laid for the development of dual theories in euclidean differential geometry of surfaces. The necessary geometric characterizations will be made to secure a purely geometric basis for these theories.

Let x denote a vector from the origin to a point on the surface whose coordinates are $x^i(u^1, u^2)$, $(i = 1, 2, 3)$. The homogeneous coordinates x^1, x^2, x^3, 1 are specialized projective coordinates which can be obtained by selecting the three cartesian coordinate planes and the plane at infinity as the faces of the tetrahedron of reference, and by choosing the unit point so that the coordinates x^i, $(i = 1, 2, 3)$, are measured with respect to equal units on the three orthogonal axes. Let z^i, $(i = 1, 2, 3)$, denote the direction cosines of the normal to S at x and let z denote the point whose coordinates are z^i, $(i = 1, 2, 3, 4)$ wherein $z^4 = 0$. The pairs of functions (x, z) are solutions of a system of differential equations of the form (1.1), wherein

$$
\begin{aligned}
&a_{ij} = c_i = e_i = 0, \qquad (i, j = 1, 2),\\
(3.1)\qquad &\binom{k}{i\ j} = \left\{ {k \atop i\ j} \right\}\\
&b_i^j = (g_{ir}d_{ik} - g_{kk}d_{ij})/g,
\end{aligned}
$$

wherein $(i, j, k, r = 1, 2)$, with $r \neq i$, $k \neq j$, in which g_{ij} and d_{ij} are the coefficients of the first and second fundamental forms, respectively, $\left\{ {k \atop i\ j} \right\}$ are the Chris-

toffel symbols of the sedond kind for the first fundamental form of the surface, and g denotes the determinant of the quantities g_{ij}.

Two fundamental limits which will be called the *first* and *second ratios of a curve C_λ at a point x* will now be introduced. Let $(y - x)^2$ denote the square of the length of the vector from x to y. The quantity

$$\rho = \lim_{X \to x} (U^2 - x)^2/(U^1 - x)^2 \tag{3.2}$$

will be called the *first ratio* of C_λ at x. Let $(y - x \mid z)$ denote the scalar product of the unit normal z and the vector from x to y. The quantity

$$\sigma = \lim_{X \to x} (U^2 - x \mid z)/(U^1 - x \mid z) \tag{3.3}$$

will be called *the second ratio of C_λ at x.*

The first ratio of C_λ at x is, clearly, given by

$$\rho = g_{22}\lambda^2/g_{11} \tag{3.4}$$

whenever the parametric net is not the minimal net.

If the minimal net is parametric, the differential equations of the surface take the form

$$\begin{gathered} \frac{\partial^2 x}{\partial u^{i2}} = \left(\frac{\partial(\log g_{12})}{\partial u^i}\right)\frac{\partial x}{\partial u^i} + d_{ii}z, \qquad \frac{\partial^2 x}{\partial u^1 \partial u^2} = d_{12}z, \\ \frac{\partial z}{\partial u^i} = -\left(d_{12}\frac{\partial x}{\partial u^i} + d_{ii}\frac{\partial x}{\partial u^j}\right)\Big/ g_{12}, \qquad (i, j = 1, 2), \quad i \neq j. \end{gathered} \tag{3.5}$$

Making use of these equations together with the expansions (1.5) $(i = 1, 2)$, we obtain

$$\begin{aligned} U^i - x = {} & \frac{\partial x}{\partial u^i}\left(\frac{du^i}{dt}\Delta t + \cdots\right) + \frac{\partial x}{\partial u^j}\left(-d_{ii}\left(\frac{du^i}{dt}\right)^3(\Delta t)^3/6g_{12} + \cdots\right) \\ & + z\left(d_{ii}\left(\frac{du^i}{dt}\right)^2(\Delta t)^2/2 + \cdots\right), \quad (i, j = 1, 2), \quad i \neq j. \end{aligned} \tag{3.6}$$

Forming the scalar squares of the right member of (3.6) and simplifying, by making use of the relations

$$\begin{gathered} \left(\frac{\partial x}{\partial u^i}\right)^2 = 0, \quad (i = 1, 2), \qquad \left(\frac{\partial x}{\partial u^1}\,\middle|\,\frac{\partial x}{\partial u^2}\right) = g_{12}, \text{ we find} \\ (U^i - x)^2 = -(d_{ii})^2\left(\frac{du^i}{dt}\Delta t\right)^4 \Big/ 12 + \cdots, \qquad (i = 1, 2), \end{gathered} \tag{3.7}$$

wherein $\dfrac{du^2}{dt} \Big/ \dfrac{du^1}{dt} = \lambda$. Consequently, if the minimal net is parametric

$$\rho = (d_{22})^2\lambda^4/(d_{11})^2. \tag{3.8}$$

If the parametric net is not the asymptotic net, it is clear that the second ratio of C_λ at x is given by

$$\sigma = d_{22}\lambda^2/d_{11}. \tag{3.9}$$

If the asymptotic net is parametric, the coefficients of the differential equations of S have the form (1.1) with $d_{11} = d_{22} = 0$ and $d_{12} \neq 0$. In the presence of these conditions the principal parts of the coefficients of z in the expansions (1.5) for $i = 1, 2$ are seen to be $\{{}_1{}^2{}_1\}\, d_{12}\left(\frac{du^1}{dt}\,\Delta t\right)^3 \Big/ 6$ and $\{{}_2{}^1{}_2\}\, d_{12}\left(\frac{du^2}{dt}\,\Delta t\right)^3 \Big/ 6$ respectively. Therefore, *if the asymptotic net is parametric, the second ratio of C_λ at x is given by*

$$\sigma = \{{}_2{}^1{}_2\}\, \lambda^3/\{{}_1{}^2{}_1\}. \tag{3.10}$$

The net $N_{\bar\lambda}$ derived from a family F_λ by the relation

$$\bar\rho(u^1, u^2) = 1/\rho(u^1, u^2), \tag{3.11}$$

wherein $\bar\rho$ and ρ are the first ratios of the curves $C_{\bar\lambda}$ and C_λ, respectively, consists of the integral curves of the differential equation

$$(du^2/du^1)^2 = (g_{11}/g_{22}\lambda)^2. \tag{3.12}$$

The directions at x of the curves of $N_{\bar\lambda}$ which pass through x, clearly separate harmonically the directions at x of the parametric curves. These directions are

$$du^2/du^1 = -g_{11}/g_{22}\lambda, \qquad du^2/du^1 = g_{11}/g_{22}\lambda. \tag{3.13}$$

It is clear that *if the parametric net is orthogonal, these are the directions of the tangents of S which bisect the angle formed by the tangents at x of the parametric curves.* The first equation of (3.13) can be obtained from equation (1.7) by replacing d_{11} and d_{22} by g_{11} and g_{22}, respectively. The tangent whose direction is defined by this equation will be called the *R_λ^*-correspondent of the tangent to C_λ at x.*

Analogously, the net N_{λ_1} derived from the family F_λ by the relation

$$\sigma_1(u^1, u^2) = 1/\sigma(u^1, u^2), \tag{3.14}$$

wherein σ_1 and σ are the second ratios of the curves C_{λ_1}, and C_λ, respectively, consists of the integral curves of the differential equation

$$(du^2/du^1)^2 = (d_{11}/d_{22}\lambda)^2. \tag{3.15}$$

The directions at x of the curves of N_{λ_1} which pass through x separate harmonically the directions at x of the parametric curves. These directions are given by

$$du^2/du^1 = -d_{11}/d_{22}\lambda, \qquad du^2/du^1 = d_{11}/d_{22}\lambda. \tag{3.16}$$

The first of these is the direction of the R_λ-correspondent of the tangent to C_λ at x.

Dual theories will be formulated in which the R_λ and R^*_λ-correspondents play fundamental roles. One of these theories may be obtained from the other by making use of the substitution

$$T = \binom{g_{ij}}{d_{ij}}, \qquad (i, j = 1, 2). \tag{3.17}$$

In these theories the geometric entities of the following pairs are the duals of each other: the asymptotic net and the minimal net, an arbitrary conjugate net and a corresponding orthogonal net, the normal curvature for a direction at a point and the radius of normal curvature for the direction, the R^*_λ- and the R_λ-correspondents of the tangent to C_λ at x.

We present a simple euclidean geometric characterization for each of the R^*_λ- and R_λ-correspondents of a tangent to C_λ at x. These characterizations are suggested by the forms of the two equations

$$du^2/du^1 = -g_{11}/g_{22}\lambda \quad \text{and} \quad du^2/du^1 = -d_{11}/d_{22}\lambda.$$

Describe along the u^1-curve through x, the arcs $x\bar{U}^1_1$ and $x\bar{U}^1_2$ of lengths equal to that of the arc xU^2, and let $x\bar{U}^1_1$ denote the one of these arcs described in the sense of the arc xU^1. Describe along the u^2-curve through x, in the sense of the arc xU^2, the arc $x\bar{U}^2_1$ of length equal to that of the arc xU^1. The u^1-curve through $\bar{U}^2_1$ intersects the u^2-curves through $\bar{U}^1_1$ and $\bar{U}^1_2$ in the points which we denote by $\bar{X}_1$ and $\bar{X}$, respectively. As X tends to x along C_λ the points $\bar{X}$ and $\bar{X}_1$ tend to x along curves whose directions at x are given by the first and second equations of (3.13), respectively. *Thus the limit of the line joining x, $\bar{X}$ as X tends to x along C_λ is the R^*_λ-correspondent of the tangent to C_λ at x.*

Let U^1_1, U^1_2 denote the points on the u^1-curve through x at distances from the tangent plane to S at x numerically equal to that from the plane to the point U^2, and let the arc xU^1_1 be described in the sense of the arc xU^1. Describe along the u^2-curve through x, in the sense of the arc xU^2, the arc xU^2_1 whose endpoint U^2_1 is at a distance from the tangent plant to S at x equal to that from the plane to the point U^1. The u^2-curves through U^1_1 and U^1_2 intersect the u^1-curve through U^2_1 in the points which we denote by X_1 and X_2, respectively. As X tends to x along C_λ, the points X_1, X_2 tend to x along curves whose directions at x are given by (3.16). *The R_λ-correspondent of the tangent to C_λ at x is the limit of the line xX_2 or the limit of the line xX_1 according as the points U^1, U^2 are on the same side or opposite sides of the tangent plane to S at x.*

We are in a position now to state metric (euclidean) definitions and theorems which are the duals of the projective definitions and theorems of Chapter I. The dual of the principal associate net of a parametric net is the net whose directions at x are the double elements of the involution $\bar{\lambda} = -g_{11}/g_{22}\lambda$. This net will be called the *metric associate net* of the parametric net. The dual of Theorem (2.1) is

THEOREM (3.1) *At a general point x of S the tangent to an arbitrary curve C_λ*

through x and the R^*_λ-correspondent of this tangent are harmonic conjugates with respect to the tangents of the metric associate net of the parametric net.

The dual of Theorem (2.2) is

THEOREM (3.2) *The net of minimal curves is the metric associate net of an arbitrary orthogonal parametric net.*

The dual of the associate conjugate net of a parametric net will be called the *associate orthogonal net of the parametric net. This net is, clearly, the unique orthogonal net whose tangents at x separate harmonically the tangents of the parametric curves. Hence, the tangents at x of the associate orthogonal net of the parametric net bisect the angle formed by the tangents at x of the parametric curves.*

The dual of the R_λ-net will be called the R^*_λ-net. The tangents of this net at a point x are the tangent to C_λ at x and the R^*_λ-correspondent of this tangent. The duals of Theorems (2.4), (2.5), (2.6) and (2.7) may now be stated as follows:

THEOREM (3.3) *Every R^*_λ-net derived in association with an arbitrary orthogonal parametric net is an orthogonal net.*

THEOREM (3.4) *Corresponding to a parametric net which does not consist of minimal curves there is a unique R^*_λ-net whose tangents at a general point x separate harmonically the tangents to the parametric curves at the point. This R^*_λ-net is the associate orthogonal net of the parametric net.*

THEOREM (3.5) *The tangents at x to the parametric curves separate harmonically the R^*_λ-correspondent of the tangent at x to a minimal curve C_λ and the tangent to the other minimal curve at x.*

THEOREM (3.6) *If the minimal net is parametric, the limit of the line joining U^1, U^2 as X tends to x along C_λ is perpendicular to the tangent to C_λ at x.*

4. Euclidean characterizations of projectively defined nets and their duals by values of the first and second ratios

If a parametric net does not consist of asymptotic curves, the equations of its principal associate net and its associate conjugate net are $d_{11}(du^1)^2 + d_{22}(du^2)^2 = 0$ and $d_{11}(du^1)^2 - d_{22}(du^2)^2 = 0$, respectively. For such a parametric net the second ratio of a curve C_λ at x is given by $\sigma = d_{22}\lambda^2/d_{11}$. Hence we have immediately

THEOREM (4.1) *If a parametric net is not the asymptotic net, its principal associate net and its associate conjugate net are such that at a general point x the second ratios of the curves of these nets are equal to -1 and $+1$, respectively.*

Dually, we have

THEOREM (4.2) *If a parametric net is not the minimal net, its metric associate net and its associate orthogonal net are such that at a general point x the first ratios of the curves of these nets are equal to -1 and $+1$, respectively.*

5. The lines of curvature; euclidean characterizations of the curves of Segre and Darboux

On an unspecialized surface there exists an infinite class of nets each of which has the property that, relative to this net as parametric the equation

$$\rho = \sigma \tag{5.1}$$

is satisfied independently of λ at a general point x of S. A net of this class will be called a duametric net. The necessary and sufficient condition that a parametric net which is not the asymptotic net nor the minimal net be a duametric net is found, by equating the right members of (3.4) and (3.9), to be

$$d_{11}g_{22} = d_{22}g_{11}\,. \tag{5.2}$$

Referred to an arbitrary parametric net, the lines of curvature are the integral curves of the equation

$$(g_{1i}d_{2j} - g_{2j}d_{1i})du^i du^j = 0, \qquad (i, j = 1, 2). \tag{5.3}$$

If the parametric net is a duametric net and the lines of curvature form a determinate net on S, equation (5.3) may be expressed in either of the forms

$$g_{11}(du^1)^2 - g_{22}(du^2)^2 = 0, \qquad d_{11}(du^1)^2 - d_{22}(du^2)^2 = 0. \tag{5.4}$$

The forms of equations (5.3) and (5.4) are such that we have immediately

THEOREM (5.1) *The associate orthogonal net of an arbitrary duametric net is the same as its dual, the associate conjugate net. This self dual net is the net of lines of curvature. Along any line of curvature* $\rho = \sigma = 1$.

Since the definition of a duametric net is self dual, *the dual of a duametric net is a duametric net.* We shall reserve for special study in §(6) a particularly important pair of *dual duametric nets.*

If the minimal curves are parametric, the first ratio of C_λ at x is defined by (3.8). The curves characterized by putting $\rho = 1$ are the integral curves of the equation

$$(d_{22})^2(du^2)^4 - (d_{11})^2(du^1)^4 = 0. \tag{5.5}$$

Equation (5.3) reduces to $d_{22}(du^2)^2 = d_{11}(du^1)^2$ on putting $g_{11} = g_{22} = 0$, $g_{12} \neq 0$. Hence the curves (5.5) consist of the lines of curvature and the integral curves of the equation

$$d_{22}(du^2)^2 + d_{11}(du^1)^2 = 0. \tag{5.6}$$

The curves (5.6) may be shown to form the orthogonal net whose tangents at a general point x bisect the angles formed by the tangents to the lines of curvature at x. For reasons which will appear in §(6) the curves of this net will be called *the lines of mean curvature.* Thus we have

THEOREM (5.2) *If the minimal curves are parametric, there are two nets which are such that at a general point x of S the first ratios of the curves of these nets are equal to* 1. *These nets consist of the lines of mean curvature and the lines of curvature of S.*

Similarly we obtain

THEOREM (5.3) *If the asymptotic curves are parametric, a curve of S is a line of curvature if and only if at each of its points its first ratio is equal to* 1.

The dual of this theorem is

THEOREM (5.4) *If the minimal curves are parametric, a curve is a line of curvature if and only if at each of its points its second ratio is equal to* 1.

The curves of Segre and of Darboux are known to be the integral curves of the equations

$$(5.7) \qquad ({}_2{}^1{}_2)(du^2)^3 - ({}_1{}^2{}_1)(du^1)^3 = 0, \qquad ({}_2{}^1{}_2)(du^2)^3 + ({}_1{}^2{}_1)(du^1)^3 = 0$$

respectively, when the asymptotic curves are parametric. When the system of coordinates for x is cartesian $({}_2{}^1{}_2) = \{{}_2{}^1{}_2\}$ and $({}_1{}^2{}_1) = \{{}_1{}^2{}_1\}$. Hence we have immediately

THEOREM (5.5) *If the asymptotic curves are parametric, a curve of S is a curve of Segre or is a curve of Darboux according as its second ratio at each of its points is equal to* 1 *or* -1, *respectively.*

6. The orthogonal R_λ-net; the lines of mean curvature and the mean conjugate curves

If the parametric net is an arbitrarily selected non-asymptotic net, the differential equation of the R_λ-net is

$$(6.1) \qquad d_{11}\lambda(du^1)^2 + (d_{22}\lambda^2 - d_{11})du^1du^2 - d_{22}\lambda(du^2)^2 = 0.$$

The necessary and sufficient condition that this net be orthogonal is that the harmonic invariant of the two forms

$$g_{ij}du^idu^j, \qquad (i, j = 1, 2) \quad \text{and} \quad d_{11}\lambda(du^1)^2 + (d_{22}\lambda^2 - d_{11})du^1du^2 - d_{22}\lambda(du^2)^2$$

vanish, that is

$$(6.2) \qquad d_{11}g_{12} + (d_{11}g_{22} - d_{22}g_{11})\lambda - d_{22}g_{12}\lambda^2 = 0.$$

Thus there are in general two families F_λ for which the R_λ-net is orthogonal. However, we observe that both of these families belong to the same R_λ-net. Hence we have

THEOREM (6.1) *Corresponding to a non-asymptotic parametric net there is, in general, a unique orthogonal R_λ-net.*

Equations (6.2) may be regarded as the differential equation of this net.

However, there is a unique parametric net which, in marked contrast with a general parametric net, is such that every associated R_λ-net will be orthogonal. Equation (6.2) shows that this case occurs only if

$$(6.3) \qquad g_{12} = 0, \qquad d_{11}g_{22} - d_{22}g_{11} = 0.$$

This parametric net is clearly the *duametric net which is orthogonal.* Thus the directions at x of the curves of this net are those perpendicular directions which separate harmonically the principal directions at x. Hence these directions bisect the angles formed by the principal directions. The associated normal curvatures for the directions are, therefore, equal. These normal curvatures are, moreover, equal to the mean of the principal normal curvatures since the sum of the normal curvatures in any two perpendicular directions is equal to the sum of the principal normal curvatures at the point. In view of this property we shall call the curves of this orthogonal duametric net *the lines of mean curvature.* We state now

THEOREM (6.2) *The unique parametric net which is such that every associated R_λ-net is orthogonal is the orthogonal duametric net. The curves of this net are the lines of mean curvature.*

The equation for the net of lines of mean curvature when referred to an arbitrary parametric net, may be found by equating the right members of

$$(6.4)\qquad \begin{aligned} 1/r &= d_{ij}du^i du^j / g_{ij} du^i du^j, \qquad (i, j = 1, 2),\\ (1/r_1 + 1/r_2)/2 &= (g_{11}d_{22} + g_{22}d_{11} - 2g_{12}d_{12})/2g \end{aligned}$$

wherein r is the radius of normal curvature of the direction du^2/du^1 and r_1, r_2 are the principal radii of normal curvature. The equation reduces to

$$(6.5)\quad (g_{11}B - g_{12}A)(du^1)^2 + (g_{11}C - g_{22}A)du^1du^2 + (g_{12}C - g_{22}B)(du^2)^2 = 0,$$

wherein A, B, C are the coefficients

$$(6.6)\quad A = 2(g_{11}d_{12} - g_{12}d_{11}), \qquad B = g_{11}d_{22} - g_{22}d_{11}, \qquad C = 2(g_{12}d_{22} - g_{22}d_{12})$$

of equation (5.3) for the lines of curvature.

If $B = 0$ and $g_{12} \neq 0$, an arbitrary non-orthogonal duametric net is parametric. For such a choice of a parametric net equations (6.2) and (5.3) both reduce to $d_{11}(du^1)^2 = d_{22}(du^2)^2$. Hence we have

THEOREM (6.3) *The orthogonal R_λ-net derived in association with an arbitrary non-orthogonal duametric net consists of the lines of curvature of S.*

If the parametric net is not duametric, we find that the necessary and sufficient conditions that the orthogonal R_λ-net consist of the lines of curvature are

$$(6.7)\qquad g_{11}d_{12} = g_{22}d_{12} = 0.$$

Hence we have

THEOREM (6.4) *If the parametric net is either the minimal net or a conjugate net and is not duametric, the associated orthogonal R_λ-net consists of the lines of curvature of S.*

The curves characterized by the properties which are the duals of those of the lines of mean curvature are found to be the mean conjugate curves. We recall that the mean conjugate curves are those for whose directions at a general point the radii of normal curvatures are equal to the mean of the principal radii of normal curvature at the point. The differential equation of the mean conjugate net is found to be

$$(6.8)\quad (d_{11}B - d_{12}A)(du^1)^2 + (d_{11}C - d_{22}A)du^1du^2 + (d_{12}C - d_{22}B)(du^2)^2 = 0,$$

wherein A, B, C are defined by (6.6). Since the net of lines of mean curvature is a duametric net, the mean conjugate net is likewise a duametric net. The mean conjugate net is parametric if and only if the equations

$$(6.9)\qquad d_{12} = 0, \qquad d_{11}g_{22} - d_{22}g_{11} = 0,$$

are satisfied identically. The duals of Theorems (6.1) to (6.4) follow:

THEOREM (6.5) *Corresponding to a non-minimal parametric net there is in general, a unique conjugate* R_λ^**-net.*

THEOREM (6.6) *The unique parametric net which is such that every associated* R_λ^**-net is conjugate is the conjugate duametric net, known heretofore as the mean conjugate net.*

THEOREM (6.7) *The conjugate* R_λ^**-net derived in association with an arbitrary non-conjugate duametric net consists of the lines of curvature of* S.

THEOREM (6.8) *If the parametric net is either the asymptotic net or an orthogonal net, the associated conjugate* R_λ^**-net consists of the lines of curvature of* S.

7. The curvatures at x of the intersection of the surface S with its tangent plane at x

Let the asymptotic curves of S be parametric and let the points x, $\dfrac{\partial x}{\partial u^1}$, $\dfrac{\partial x}{\partial u^2}$ be the vertices of a *local triangle of reference at* x, with a unit point chosen so that any point whose general coordinates are given by an expression of the form

$$x_1 x + x_2 \frac{\partial x}{\partial u^1} + x_3 \frac{\partial x}{\partial u^2}$$

shall have local coordinates x_1, x_2, x_3. Let us introduce non-homogeneous coordinates by the definitions

$$\xi = x_2/x_1, \qquad \eta = x_3/x_1. \tag{7.1}$$

The intersection of S with its tangent plane at an ordinary point x is a curve C which has a node at x, the nodal tangents being the asymptotic tangents of S at x. In a neighborhood of x the equation for the branch of C which is tangent to the u^1-tangent is given by the power series development[4]

$$\eta = ({}_1{}^2{}_1)(\xi^2/3 - \varphi\xi^3/12) + \cdots, \tag{7.2}$$

wherein $\varphi = ({}_1{}^1{}_1) + \dfrac{\partial \log ({}_1{}^2{}_1)}{\partial u^1}$. Similarly, we have along the branch of C which is tangent to the u^2-tangent the development

$$\xi = ({}_2{}^1{}_2)(\eta^2/3 - \psi\eta^3/12) + \cdots, \tag{7.3}$$

wherein $\psi = ({}_2{}^2{}_2) + \dfrac{\partial \log ({}_2{}^1{}_2)}{\partial u^2}$.

Let the coordinates x be the homogeneous cartesian coordinates x_1, x_2, x_3, 1. Then $({}_i{}^k{}_j)$, $(i, j, k = 1, 2)$, are defined by (3.1). Since $x_4 = 1$, the points $\dfrac{\partial x}{\partial u^1}$ and $\dfrac{\partial x}{\partial u^2}$ are the points at infinity on the u^1- and u^2-tangents of S at x. Let us alter the unit point so that new local coordinates $(\bar\xi, \bar\eta)$ are measured with respect to equal units on the axes. Since $\dfrac{\partial x}{\partial u^1} \Big/ (g_{11})^{\frac{1}{2}}$ and $\dfrac{\partial x}{\partial u^2} \Big/ (g_{22})^{\frac{1}{2}}$ are unit vectors in

[4] Cf. P. D. G. page 73.

the directions of the u^1- and u^2-tangents to S at x, we must have the identity

$$(7.4)\qquad x + \xi \frac{\partial x}{\partial u^1} + \eta \frac{\partial x}{\partial u^2} \equiv x + \bar{\xi} \frac{\partial x}{\partial u^1} \Big/ (g_{11})^{\frac{1}{2}} + \bar{\eta} \frac{\partial x}{\partial u^2} \Big/ (g_{22})^{\frac{1}{2}}.$$

Hence, we have the relations

$$(7.5)\qquad \xi = \bar{\xi}/(g_{11})^{\frac{1}{2}}, \qquad \eta = \bar{\eta}/(g_{22})^{\frac{1}{2}}.$$

In terms of $\bar{\xi}$, $\bar{\eta}$ equations (7.2) and (7.3) take the respective forms

$$(7.6)\qquad \bar{\eta} = \left\{ {}_{1}{}^{2}{}_{1} \right\} (g_{22})^{\frac{1}{2}} \bar{\xi}^2/3g_{11} + \cdots ,$$

$$(7.7)\qquad \bar{\xi} = \left\{ {}_{2}{}^{1}{}_{2} \right\} (g_{11})^{\frac{1}{2}} \bar{\eta}^2/3g_{22} + \cdots .$$

With respect to this oblique coordinate system defined by (7.5) the equation of a circle whose center is the point (h, k) and whose radius is r is found to be

$$(7.8)\qquad (\bar{\xi} - h)^2 + (\bar{\eta} - k)^2 + 2(\bar{\xi} - h)(\bar{\eta} - k) g_{12}/(g_{11}g_{22})^{\frac{1}{2}} = r^2.$$

Demanding that this equation be satisfied by the series (7.6) identically in $\bar{\xi}$ as far as the terms of the second degree, we obtain the conditions that the circle (7.8) osculate at x the branch of C which is tangent to the u^1-tangent at x

$$(7.9)\qquad h = -3(g_{11})^{\frac{3}{2}} g_{12}/2\left\{ {}_{1}{}^{2}{}_{1} \right\} g, \quad k = 3(g_{11})^2 (g_{22})^{\frac{1}{2}}/2\left\{ {}_{1}{}^{2}{}_{1} \right\} g, \quad r^2 = 9(g_{11})^3/4\left\{ {}_{1}{}^{2}{}_{1} \right\}^2 g.$$

The substitution

$$T = \begin{pmatrix} g_{11} & g_{12} & \left\{ {}_{1}{}^{2}{}_{1} \right\} & h & \bar{\xi} \\ g_{22} & g_{21} & \left\{ {}_{2}{}^{1}{}_{2} \right\} & k & \bar{\eta} \end{pmatrix}$$

transforms equation (7.2) into equation (7.3) and equation (7.8) with h, k, r defined by (7.9) into an equation of form (7.8) with k, h, r defined by

$$(7.10)\qquad k = -3(g_{22})^{\frac{3}{2}} g_{12}/2\left\{ {}_{2}{}^{1}{}_{2} \right\} g, \qquad h = 3(g_{22})^2 (g_{11})^{\frac{1}{2}}/2\left\{ {}_{2}{}^{1}{}_{2} \right\} g,$$
$$r^2 = 9(g_{22})^3/4\left\{ {}_{2}{}^{1}{}_{2} \right\}^2 g.$$

Hence, *the circle* (7.8) *with* h, k, r *defined by* (7.10) *osculates at* x *the branch of* C *which is tangent to the* u^2*-tangent at* x.

The equation of the radical axis of the two circles which osculate C at x is readily found to be

$$(7.11)\qquad \left\{ {}_{1}{}^{2}{}_{1} \right\} (g_{22})^{\frac{1}{2}} \bar{\xi} - \left\{ {}_{2}{}^{1}{}_{2} \right\} (g_{11})^{\frac{1}{2}} \bar{\eta} = 0.$$

In terms of the original local coordinates ξ, η this equation is

$$(7.12)\qquad \left\{ {}_{1}{}^{2}{}_{1} \right\} g_{22} \xi - \left\{ {}_{2}{}^{1}{}_{2} \right\} g_{11} \eta = 0.$$

The direction of this line is given by

$$\eta/\xi = g_{22}\left\{ {}_{1}{}^{2}{}_{1} \right\}/g_{11}\left\{ {}_{2}{}^{1}{}_{2} \right\}.$$

The direction of the R_λ-correspondent of the tangent at x in either principal direction is given by

$$du^2/du^1 = -g_{22}\{{}_1{}^2{}_1\}/g_{11}\{{}_2{}^1{}_2\}$$

Hence we have

THEOREM (7.1) *If the asymptotic curves are the parametric curves of S, the R_λ-correspondent of the tangent at x in either principal direction and the radical axis of the osculating circles of C at x are conjugate tangents.*

On observing the forms (3.4) and (3.10) for the first and second ratios of a curve C_λ at x we find that the only non-asymptotic direction λ for which these ratios are equal is that defined by

$$\lambda = g_{22}\{{}_1{}^2{}_1\}/g_{11}\{{}_2{}^1{}_2\}. \tag{7.13}$$

Hence we have

THEOREM (7.2) *If a curve C_λ is not tangent to an asymptotic curve at x, its tangent line at x is the radical axis of the osculating circles of C at x if and only if first and second ratios of C_λ at x are equal.*

Let K_1K_2 denote the curvatures of the circles which osculate at x the branches of C which are tangent to the u^1- and u^2-tangents respectively. These curvatures will be called the *curvatures of tangential section of S at x.* We have in view of (7.9) and (7.10)

$$K_1 = 2e_1\{{}_1{}^2{}_1\}g^{1/2}/3(g_{11})^{3/2}, \qquad k_2 = 2e_2\{{}_2{}^1{}_2\}g^{1/2}/3(g_{22})^{3/2}, \tag{7.14}$$

wherein $g^{1/2} = (g_{11}g_{22} - g_{12}^2)^{1/2}$ and $e_1 = \pm 1$, $e_2 = \pm 1$, the signs of e_1 and e_2 being chosen so that K_1 and K_2 are positive. These curvatures are equal if and only if

$$(g_{11}/g_{22})^{1/2} = e(\{{}_1{}^2{}_1\}/\{{}_2{}^1{}_2\})^{1/3}, \tag{7.15}$$

with $e = \pm 1$ the sign being selected so that the right member is positive. The left member of (7.15) is a principal direction and the right member is a direction of Segre or a direction of Darboux according as e is $+1$ or -1 respectively. Substituting in (7.13) the expression obtained for g_{22}/g_{11} from (7.15), we find that (7.13) becomes

$$\lambda = (\{{}_1{}^2{}_1\}/\{{}_2{}^1{}_2\})^{1/3}. \tag{7.16}$$

Thus we find that under conditions (7.15) the direction defined by (7.13) is a direction of Segre and also, in consequence of (7.15), a principal direction. Moreover, if the direction defined by (7.13) is a direction of Segre, equations (7.15) follow. We may therefore state

THEOREM (7.3) *The necessary and sufficient condition that the curvatures of tangential section of S at x be equal is that the direction of the radical axis of the osculating circles of C at x be a direction of Segre. The condition is, moreover,*

necessary and sufficient that this direction of Segre and its conjugate direction of Darboux be principal directions at x.

In view of Theorems (5.3), (5.5) and (7.3) we state in conclusion

COROLLARY (7.4) *If the asymptotic curves are parametric, the one parameter family F_λ of curves enveloped by the radical axis of the osculating circles of C at x as x varies over S has the property that the first and second ratios of C_λ at x are both equal to* 1 *if and only if the curvatures of tangential section of S at x are equal.*

PRINCETON UNIVERSITY
UNIVERSITY OF KANSAS

Annals of Mathematics
Vol. 44, No. 2, April, 1943

ON FAMILIES OF MUTUALLY EXCLUSIVE SETS

By P. Erdös and A. Tarski

(Received August 11, 1942)

In this paper we shall be concerned with a certain particular problem from the general theory of sets, namely with the problem of the existence of families of mutually exclusive sets with a maximal power. It will turn out—in a rather unexpected way—that the solution of these problems essentially involves the notion of the so-called "inaccessible numbers." In this connection we shall make some general remarks regarding inaccessible numbers in the last section of our paper.

§1. FORMULATION OF THE PROBLEM. TERMINOLOGY[1]

The problem in which we are interested can be stated as follows: Is it true that every field $\mathfrak{F}$ of sets contains a family of mutually exclusive sets with a maximal power, i.e. a family $\mathfrak{G}$ whose cardinal number is not smaller than the cardinal number of any other family $\mathfrak{H}$ of mutually exclusive sets contained in $\mathfrak{F}$.

By a field of sets we understand here as usual a family $\mathfrak{F}$ of sets which together with every two sets X and Y contains also their union $X \cup Y$ and their difference $X - Y$ (i.e. the set of those elements of X which do not belong to Y) among its elements. A family $\mathfrak{G}$ is called a family of mutually exclusive sets if no set X of $\mathfrak{G}$ is empty and if any two different sets of $\mathfrak{G}$ have an empty intersection.

A similar problem can be formulated for other families e.g. for rings of sets, i.e. for families which together with any two sets X and Y also contain their union $X \cup Y$ and their intersection $X \cap Y$ among their elements. We obtain an especially interesting particular case of this problem by referring it to the ring of open sets of a topological space S with power $2^{\aleph_0}$.

It turns out that the solution of our problem is in general positive; however it is negative in certain exceptional cases. To examine the problem thoroughly we must first subject it to a certain transformation by using some notions from the arithmetic of cardinal numbers.

We shall denote the cardinal number (or power) of a set S by $\mathfrak{c}(S)$.

A cardinal number $\mathfrak{n}$ is called a limit number if $\mathfrak{n} \neq 0$ and if among the cardinal numbers $\mathfrak{x} < \mathfrak{n}$ there is no largest one. The number $\mathfrak{n}$ is called *singular* if it can be expressed as a sum of less than $\mathfrak{n}$ numbers $\mathfrak{m}$, each of which is smaller than $\mathfrak{n}$.

---

[1] For the concepts and results of the general theory of sets, which are applied in this paper, see Hausdorff, Mengenlehre: however as regards the concept of an inaccessible number cf. Tarski, Über unerreichbare Kardinalzahlen, Fund. Math. Vol. 30 (1938) p. 68–89. For the concepts and results from the theory of partially ordered sets, lattices, Boolean algebras, etc., see G. Birkhoff, Lattice theory. For topological concepts see Kuratowski, Topologie I.

If such a representation is impossible the number $\mathfrak{n}$ is called *regular*. Regular limit numbers are also referred to as "inaccessible" or "weakly inaccessible" numbers.

As is well known, every limit number is an infinite number, and every singular infinite number is a limit number. The problem of the existence of regular limit numbers $> \aleph_0$ is thus far unsolved, and presumably will never be solved on the basis of the axiom systems upon which the general theory of sets is constructed at present. At any rate the existence of the numbers in question cannot be derived from these axiom systems provided they are consistent; on the other hand it seems highly improbable that these systems cease to be consistent if we enrich them by adding new existential axioms which secure the existence of the inaccessible numbers.

$\mathfrak{F}$ being a family of sets let us denote by $\mathfrak{d}(\mathfrak{F})$ the smallest cardinal number which is $> \mathfrak{c}(\mathfrak{G})$ for every family $\mathfrak{G}$ of mutually exclusive sets contained in F. If $\mathfrak{d}(\mathfrak{F})$ is not a limit number, the family F obviously contains a subfamily $\mathfrak{G}$ of mutually exclusive sets with a maximal power. Thus our problem reduces now to the following one:

$\mathfrak{n}$ being a limit number is it true that for every field (or ring) of sets we have $\mathfrak{d}(\mathfrak{F}) \neq \mathfrak{n}$?

We shall show that the solution of this problem depends on the properties of the number $\mathfrak{n}$: the answer is affimative if $\mathfrak{n}$ either $= \aleph_0$ or is a regular number (Theorem 1), is negative only for the hypothetical regular limit numbers $> \aleph_0$ (Theorem 2). If in particular the problem is applied to the ring of all open sets of a topological space[2] with $\mathfrak{c}(S) = 2^{\aleph_0}$, then its positive solution proves to be equivalent with the statement that there is no inaccessible number $> \aleph_0$ and $\leqq 2^{\aleph_0}$ (Corollary 3).

In order to formulate the positive part of our result in as general form as possible, we shall use the terminology of partially ordered sets.

Let S be an arbitrary set which is partially ordered by the binary relation $\leqq$. If x is an element of S, we write $S(x)$ to denote the partially ordered set of all elements $y \epsilon S$ which are $\leqq x$. The symbol $\wedge$ will denote a null element of S, i.e. an element x such that $x \leqq y$ for every $y \epsilon S$. Two elements y and z of S are called disjoint if $y \neq \wedge$, $z \neq \wedge$ and if for every $x \epsilon S$ the formulas $x \leqq y$ and $x \leqq z$ imply $x = \wedge$. We do not here assume that the partially ordered set necessarily contains a null element. In fact without loss of generality we could confine ourselves to the consideration of sets which do not contain such elements; and in this case we could simply say that two elements y and z are called disjoint if there is no element x such that $x \leqq y$ and $x \leqq z$.

A subset T of a partially ordered set S such that every element of T is $\neq \wedge$ and every two different elements of T are disjoint is called a set of mutually

[2] By topological spaces we mean here the spaces with the closure operation satisfying the axioms I–III of Kuratowski (op. cit. p. 77). However the space which will be constructed in the proof of Corollary 2 will also satisfy axiom IV (normality) (pp. 95–101, ibid.).

exclusive elements. Again we denote by $\mathfrak{d}(\mathfrak{S})$ the smallest cardinal number $> \mathfrak{c}(T)$ for every $T \subseteq S$ of sets of mutually exclusive elements; moreover we write for every element $x \epsilon T$

$$\mathfrak{d}(x) = \mathfrak{d}(\mathfrak{S}(x))$$

In view of this formula $\mathfrak{d}$ constitutes an example of a function f which correlates with every element of a partially ordered set a cardinal number $f(x)$. This function is obviously increasing, for we have

$$\mathfrak{d}(x) \leqq \mathfrak{d}(y)$$

for every two elements x and y such that $x \leqq y$. Many other examples of this kind of increasing functions are also known; e.g. the function c defined for every $x \epsilon S$ by the formula

$$\mathfrak{c}(x) = \mathfrak{c}(S(x)).$$

Still another example is constituted by the function g defined in the following way: for every $x \epsilon S$, $g(x)$ is the smallest cardinal number $\mathfrak{n}$ such that there is a basis B of the set $S(x)$ with power $c(B) = \mathfrak{n}$; by a basis we here understand a set $B \subseteq S(x)$ such that every element of $S(x)$ is the union (the least upper bound) of elements of B. To every increasing function f of the kind considered there corresponds a certain notion of *homogeneity* of partially ordered sets. We say generally that an element x of a partially ordered set S is homogeneous with respect to an increasing function f, which is defined over the set S and assumes cardinal numbers as values, or simply that x is f-homogeneous, if $x \neq \Lambda$ and if $f(x) = f(y)$ for every element $y \epsilon S$ such that $y \neq \Lambda$ and $y \leqq x$. If the set S contains a unit element u i.e. an element x such that $y \leqq x$ for every $y \epsilon S$), and if u is f-homogeneous, the whole set S is called f-homogeneous.

SOLUTION OF THE PROBLEM

We shall begin with two simple lemmas concerning f-homogeneous elements

LEMMA 1. *Let S be a partially ordered set, and f an increasing function which correlates with every element $x \epsilon S$ a cardinal number $f(x)$. Then for every element $x \neq \Lambda$ there exists an f-homogeneous element $y \leqq x$.*

PROOF. Consider all the cardinal numbers $f(y)$ correlated with the elements $y \leqq x, y \neq \Lambda$. Among these cardinal numbers there certainly exists a smallest, say $\mathfrak{n}$ (by the well ordering theorem); and it is easily seen that every element such that

$$y \leqq x, \qquad f(y) = \mathfrak{n}$$

is f-homogeneous.

LEMMA 2. *Under the hypothesis of* LEMMA 1. *there exists a set $T \subseteq S$ of mutually exclusive f-homogeneous elements such that no element of S is disjoint with all elements of T.*

PROOF. It can be easily shown (e.g. with the help of well ordering) that there exists a maximal set T of mutually exclusive f-homogeneous elements of S; i.e., a set T of mutually exclusive f-homogeneous elements of S which is not a proper subset of any other set with the same property. Hence by LEMMA 1 it follows that no element of S—whether homogeneous or not—is disjoint with every element of T, q.e.d.

As an immediate consequence of LEMMA 2 we obtain the following theorem which, however, will not be applied in this paper.

Let B be a Boolean algebra, and f an increasing function which correlates with every element x of B a cardinal number $f(x)$. Then every element of B—and, in particular, the unit element—can be represented as the union of mutually exclusive f-homogeneous elements of B; and therefore B is isomorphic with a direct sum of f-homogeneous Boolean algebras.

The following three lemmas will lead us directly to THEOREM 1, which is one of the main results of this paper.

LEMMA 3. *If S is a partially ordered set and $\mathfrak{d}(S)$ is a limit number, then S contains a $\mathfrak{d}$-homogeneous element x with $\mathfrak{d}(x) = \mathfrak{d}(S)$.*

PROOF. Assume that, on the contrary, S does not contain a $\mathfrak{d}$-homogeneous element x with $\mathfrak{d}(x) = \mathfrak{d}(S)$. By applying LEMMA 2 to the function $f = \mathfrak{d}$ we obtain a set $T \subseteq S$ of mutually exclusive $\mathfrak{d}$-homogeneous elements, with the property that no element of S is disjoint with every element of T. According to our assumption we have:

$$\mathfrak{d}(t) < \mathfrak{d}(S) \quad \text{for every element} \quad t \in T; \tag{1}$$

moreover, the definition of $\mathfrak{d}$ implies:

$$\mathfrak{c}(T) < \mathfrak{d}(S) \tag{2}$$

Since $\mathfrak{d}(S)$ is an infinite cardinal number, we have

$$(\mathfrak{d}(S))^2 = \mathfrak{d}(S); \tag{3}$$

hence, by (1) and (2), we obtain:

$$\sum_{t \in T} \mathfrak{d}(t) \leqq \mathfrak{c}(T) \cdot \mathfrak{d}(S) \leqq (\mathfrak{d}(S))^2 = \mathfrak{d}(S). \tag{4}$$

We want now to show that in the latter formula '$\leqq$' may be replaced by '$=$'. In fact, consider an arbitrary set $U \subseteq S$ of mutually exclusive elements. As was mentioned before, no element of U can be disjoint with every element of T. Hence (by using the axiom of choice) we can correlate with every element $u \in U$ first an element $t_u \in T$, and then an element $v_u \in S$ such that $v_u \neq \Lambda$, $v_u \leqq u$, and $v_u \leqq t_u$. Let V be the set of all these elements v_u. It can easily be seen that the correspondence between the elements of U and those of V is one-to-one, and that therefore the sets U and V have the same power. If, on the other hand, we denote by V_t (where t is a given element of T) the set of all those elements v_u which are $\leqq t$, we see at once that V is the union of all these sets V_t;

$$V = \bigcup V_t$$

and that

$$\mathfrak{c}(V_t) < \mathfrak{d}(t) \quad \text{for every} \quad t \in T.$$

Hence

$$\mathfrak{c}(U) = \mathfrak{c}(V) \leqq \sum_{t \in T} \mathfrak{d}(t).$$

Since the latter formula holds for every set $U \subseteq S$ of mutually exclusive elements, we infer from the definition of $\mathfrak{d}$ that either

$$\mathfrak{d}(S) \leqq \sum_{t \in T} \mathfrak{d}(t) \tag{5}$$

or else $\mathfrak{d}(S)$ is the cardinal number which immediately follows $\sum_{t \in T} \mathfrak{d}(t)$. However, the second alternative is excluded, $\mathfrak{d}(S)$ being by hypothesis a limit number, and therefore (5) holds. The formulas (4) and (5) give at once:

$$\sum_{t \in T} \mathfrak{d}(t) = \mathfrak{d}(S). \tag{6}$$

From (1), (2), (3), and (6) it follows that for every cardinal number $\mathfrak{x} < \mathfrak{d}(S)$ there exists an element $t \in T$ such that $\mathfrak{x} < \mathfrak{d}(t)$. For if we had:

$$\mathfrak{d}(S) > \mathfrak{x} \geqq \mathfrak{d}(t) \quad \textit{for every} \quad t \in T$$

we should have, by (1), (2), and (3),

$$\sum_{t \in T} \mathfrak{d}(t) \leqq c(T)\mathfrak{x} < (\mathfrak{d}(S))^2 = \mathfrak{d}(S)$$

which obviously contradicts (6). Hence we can easily construct (with the help of the well ordering theorem) a well ordered transfinite sequence of elements $t_0, t_1, \cdots, t_\xi, \cdots \in T$ of an ordinal type τ which satisfy the following conditions:

$$\mathfrak{d}(t_{\xi_1}) < \mathfrak{d}(t_{\xi_2}) \quad \text{for} \quad \xi_1 < \xi_2 < \tau, \qquad \tau \text{ being a limit ordinal number,} \tag{7}$$

and

$$\sum_{\xi < \tau} \mathfrak{d}(t_\xi) = \mathfrak{d}(S). \tag{8}$$

Consider an arbitrary ordinal number $\xi < \tau$. By (7) we have

$$\mathfrak{d}(t_\xi) < \mathfrak{d}(t_{\xi+1})$$

Hence by virtue of the definition of $\mathfrak{d}$, there exists a set $W_\xi \subseteq S(t_{\xi+1})$ of mutually exclusive elements with power:

$$\mathfrak{c}(W_\xi) = \mathfrak{d}(t_\xi) \quad \text{for every} \quad \xi < \tau. \tag{9}$$

Putting

$$W = \bigcup_{\xi < \tau} W_\xi \tag{10}$$

we easily see that W is a set of mutually exclusive elements of S, the sets $W_0, W_1, \cdots, W_\xi, \cdots$ being mutually exclusive. We have moreover, by (8), (9), and (10),

$$\mathfrak{c}(W) = \mathfrak{d}(S).$$

In this way we have arrived at a contradiction; for $\mathfrak{d}(S)$ is by definition $> \mathfrak{c}(X)$ for every set $X \subseteq S$ of mutually exclusive elements. Thus we must reject our original supposition, and assume that S has a d-homogeneous element x with $\mathfrak{d}(x) = \mathfrak{d}(S)$, q.e.d.

LEMMA 4. *If x is a $\mathfrak{d}$-homogeneous element of a partially ordered set S, then $\mathfrak{d}(x) \neq \aleph_0$.*

PROOF. Assume $\mathfrak{d}(x) = \aleph_0$. By the definition of $\mathfrak{d}$ there exist two disjoint elements x_1 and x_2 which are $\leqq x$. Since x is $\mathfrak{d}$-homogeneous, we have further $\mathfrak{d}(x_2) = \aleph_0$, and therefore there exist two disjoint elements, $x_{2,1}$ and $x_{2,2}$ which are $\leqq x_2 \leqq x$. By continuing this procedure indefinitely we obtain (with the help of the axiom of choice) an infinite sequence of mutually exclusive elements $x_1, x_{2,1}, x_{2,2,1}, \cdots$ which are all $\leqq x$; but this clearly contradicts our assumption. Hence $\mathfrak{d}(x) \neq \aleph_0$, q.e.d.

LEMMA 5. *If x is a $\mathfrak{d}$-homogeneous element of a partially ordered set, then $\mathfrak{d}(x)$ is not a singular limit number.*

PROOF. Assume, on the contrary, that $\mathfrak{d}(x)$ is a singular limit number. Thus $\mathfrak{d}(x)$ can be represented in the form:

$$\mathfrak{d}(x) = \sum_{i \in C} \mathfrak{m}_i \tag{11}$$

where C is a certain set of power $< \mathfrak{d}(x)$, and every number $\mathfrak{m}_i$ is also $< \mathfrak{d}(x)$. Since $\mathfrak{c}(C) < \mathfrak{d}(x)$, we can correlate with every element $i \in C$ an element $x_i \leqq x$ in such a way that any two elements x_{i_1} and x_{i_2} $(i_1 \neq i_2)$ are disjoint. The element x being $\mathfrak{d}$-homogeneous, we have for every element x_i :

$$\mathfrak{d}(x_i) = \mathfrak{d}(x) > \mathfrak{m}_i ;$$

consequently we can correlate with every element x_i a set $T_i \subseteq S(x_i)$ of mutually exclusive elements with power $\mathfrak{c}(T_i) = \mathfrak{m}_i$. Hence in view of (11) it is easily seen that the set T defined by means of the formula

$$T = \bigcup_{i \in C} T_i$$

is a set of mutually exclusive elements of $S(x)$ with power

$$\mathfrak{c}(T) = \sum_{i \in C} \mathfrak{M}_i = \mathfrak{d}(x).$$

But this is impossible, since $\mathfrak{d}(x)$ must be by definition $> \mathfrak{c}(T)$. Thus we must assume that $\mathfrak{d}(x)$ is not a singular limit number, q.e.d.

Lemmas 3, 4, and 5 imply directly

THEOREM 1. *If $\mathfrak{n}$ is either equal to $\aleph_0$ or is a singular limit number then there is no partially ordered set S such that $\mathfrak{d}(S) = \mathfrak{n}$.*

REMARK. Theorem 1 applies directly to various special partially ordered sets, e.g., to lattices or Boolean algebras. It can also be applied to an arbitrary family $\mathfrak{F}$ of sets, for every such family is partially ordered by the relation of inclusion $\subseteq$. It should be noticed, however, that two sets of a family $\mathfrak{F}$ which are disjoint from the point of view of the theory of partially ordered sets are not necessarily disjoint in the usual set-theoretic meaning. On the other hand, it is easily seen that the two meanings of the notion of disjointness coincide if the family $\mathfrak{F}$ contains the empty set among its elements and if, with any two sets X and Y belonging to $\mathfrak{F}$ their intersection $X \cap Y$ also belongs to $\mathfrak{F}$. Thus Theorem 1 applies literally to every field of sets, and even to every ring of sets which contains the empty set, e.g., to the ring of all open sets of a topological space; and it can be very easily shown that the theorem holds for an arbitrary ring of sets, even if it does not contain the empty set (for in this case the ring does not contain any two sets which are disjoint in the set-theoretic sense).

THEOREM 2. *If* $\mathfrak{n}$ *is a regular cardinal number* $> \aleph_0$, *then for every set* S *of power* $\mathfrak{n}$ *there exists a field* $\mathfrak{F}$ *of subsets of* S *such that* $\mathfrak{c}(F) = \mathfrak{d}(\mathfrak{F}) = n$.

PROOF. We could assume that $\mathfrak{n}$ is a limit number, for otherwise the proof presents no difficulty; however, no use will be made here of this assumption.

We shall first prove the theorem for a particular set N of power $\mathfrak{n}$, which will be defined as follows. Let us write for every ordinal number α:

(1) $\mathfrak{c}(\alpha) =$ the power of the set of all ordinal numbers $\xi < \alpha$. By the well-ordering theorem there exists an ordinal number ν such that

(2) $\mathfrak{c}(\nu) = \mathfrak{n}$, while $\mathfrak{c}(\xi) < \mathfrak{n}$ for every number $\xi < \nu$.

(3). $N =$ the set of all transfinite sequences σ of ordinal numbers $\sigma_0, \sigma_1, \cdots,$ which satisfy the following conditions:

(i) $\sigma_\xi \leqq \xi$ for every number $\xi < \nu$;

(ii) there are only finitely many numbers $\xi < \nu$ such that $\sigma_\xi \neq 0$.

Since

(4). $n = n^0 + n^1 + \cdots + n^k + \cdots \qquad (k < \aleph_0)$,

it is easily seen from (1), (2), and (3) that N has in fact power $\mathfrak{n}$.

We are now going to correlate to every number $\xi < \nu$ a family $\mathfrak{H}$ of subsets X of N so as to satisfy the following conditions:

(5). $\mathfrak{H}_\xi$ is a family of mutually exclusive sets, and $\bigcup_{X \in \mathfrak{H}_\xi} X = N$

(6). $\mathfrak{c}(\mathfrak{H}_\xi) = \mathfrak{c}(\xi + 1) < \mathfrak{n}$;

(7). if $\xi_1, \xi_2, \cdots \xi_n$ is any finite sequence of distinct ordinal numbers $< \nu$, and $X_1, X_2, \cdots, X_n$ any finite sequence of sets such that $X_1 \in \mathfrak{H}_{\xi_1}$, $X_2 \in \mathfrak{H}_{\xi_2}, \cdots, X_n \in \mathfrak{H}_{\xi_n}$, then the intersection $X_1 \cap \cdots \cap X_n$ is not empty. To obtain such families $\mathfrak{H}_\xi$ we put:

(8). $N_{\xi,\eta} =$ the set of all sequences $\sigma \in N$ such that $\sigma_\xi = \eta$, $\eta \leqq \xi < \nu$

(9). $\mathfrak{H}_\xi =$ the family of all sets $N_{\xi,0}, N_{\xi,1} \cdots N_{\xi,\eta} \cdots$ where $\eta \leqq \xi < \nu$.

The proof that the families $\mathfrak{H}_\xi$ thus defined satisfy the conditions (5), (6), and (7) does not present any difficulties.

Finally we construct the field $\mathfrak{F}$ by putting

(10). $\mathfrak{H} = \bigcup_{\xi < \nu} \mathfrak{H}_\xi$

(11). $\mathfrak{F}$ = the smallest field of sets which contains all the sets of $\mathfrak{H}$; or, in other words,: $\mathfrak{F}$ = the family of all sets which are finite unions of finite intersections of sets $X \epsilon \mathfrak{H}$ and their complements $N - X$.

We shall prove that $\mathfrak{F}$ satisfies the conclusion of our theorem. If $\mathfrak{n}$ is an infinite number, it is easily seen from (2), (6), and (10) that the family $\mathfrak{H}$ has power $\mathfrak{n}$. Hence by (4) and (11) it follows that $\mathfrak{F}$ has also power $\mathfrak{n}$. Furthermore, (2), (5), (6), (10), and (11) imply that, for every number $\mathfrak{r} < \mathfrak{n}$, $\mathfrak{F}$ does contain $\mathfrak{r}$ mutually exclusive sets. Hence we have

(12). $\mathfrak{c}(\mathfrak{F}) = \mathfrak{n} \leqq \mathfrak{d}(\mathfrak{F})$.

It remains to show that every family $\mathfrak{G} \subseteq \mathfrak{F}$ of mutually exclusive sets has a power $<n$. We shall show it first for families of a rather special character.

Let us agree to say that a set $X \epsilon \mathfrak{F}$ is of the l^{th} order (where l is any positive integer) if it is not empty and can be represented as an intersection of l different sets of the family $\mathfrak{H}$. We are going to establish certain simple properties of sets of the l^{th} order.

(13). Every set X of the l^{th} order can be represented uniquely in the form

$$X = X_1 \cap \cdots \cap X_l$$

where $X_1 \epsilon \mathfrak{H}_{\xi_1}, \cdots, X_l \epsilon \mathfrak{H}_{\xi_l}$, and $\xi_i < \xi_k < \nu$ for $1 \leqq i < k \leqq l$.

In fact, the possibility of such a representation follows directly from the definition of the sets of the l^{th} order; two different sets X_i and X_k cannot belong to the same family $\mathfrak{H}$, for by (5) the set $X \subseteq X_i \cap X_k$ would be then empty. Assume that the set X has two representations of this kind:

$$X = X_1 \cap \cdots \cap X_l = Y_1 \cap \cdots \cap Y_l$$

where $X_i \epsilon \mathfrak{H}_{\xi_i}$, $Y_i \epsilon H_{\eta_i}$, $\xi_i < \xi_k < \nu$, and $\eta_i < \eta_k < \nu$ for $1 \leqq i < k \leqq l$. If these representations are different, at least one of the sets $X_1, \cdots, X_l$, let us say X_i, cannot occur among the sets $Y_1, \cdots, Y_l$; and similarly a certain set Y_j cannot occur among the sets $X_1, \cdots, X_l$. Hence the number ξ_i must be different from each of the numbers $\eta_1, \cdots, \eta_l$. For, if we had $\xi_i = \eta_k$ $(1 \leqq k \leqq l)$, the sets X_i and Y_k $(X_i \neq Y_k)$ would belong to the same class $\mathfrak{H}_{\xi_i} = \mathfrak{H}_{\eta_k}$; and therefore by (5) the set $X \subseteq X_i \cap Y_k$ would be empty. For the same reason the number η_j must be different from each of the numbers $\xi_1, \cdots, \xi_l$. Thus, in particular, $\xi_i \neq \eta_j$, and therefore at least one of these two numbers is $\neq 0$. Assume, e.g., $\xi_i \neq 0$. By (6) there is a set $X'_i \epsilon \mathfrak{H}_{\xi_i}$ which is different from X_i. By (5) the sets X_i and X'_i are disjoint; consequently the intersection

$$X'_i \cap X = X'_i \cap X_1 \cap \cdots \cap X_l$$

is empty. Hence the intersection

$$X'_i \cap X = X'_i \cap Y_1 \cap \cdots \cap Y_l$$

must also be empty; but this clearly contradicts (7), all the numbers $\xi_i, \eta_1, \eta_2, \cdots \eta_l$ being distinct. Thus the two representations of X cannot be different.

In what follows we shall refer to the sets $X_1, \cdots, X_l$ occurring in the representation (13) of a set X as the *factors* of X.

(14). In order that two sets X and Y of the l^{th} order be disjoint it is necessary and sufficient that they have two factors X_i and Y_i which are different, but belong to the same family $\mathfrak{H}$.

This follows directly from (5), (7), and (13).

(15). If two disjoint sets X and Y of the $(l + 1)^{\text{th}}$ order have a common factor $X_i = Y_j$, and if X^* and Y^* are the intersections of the remaining factors of X and Y respectively, then X^* and Y^* are disjoint sets of the l^{th} order.

This can be easily obtained from (13) and (14).

Now we can prove by induction with respect to l:

(16). Every family $\mathfrak{G} \subseteq \mathfrak{F}$ of mutually exclusive sets of the l^{th} order has power $< \mathfrak{n}$.

(16) is clearly true for $l = 1$. In fact, in this case $\mathfrak{G}$ is contained in the family

$$\mathfrak{H} = \bigcup_{\xi < \nu} \mathfrak{H}_\xi .$$

If there were two sets X_1 and X_2 of $\mathfrak{G}$ which belonged to two different families $\mathfrak{H}_{\xi_1}$ and $\mathfrak{H}_{\xi_2}$, they would not be disjoint, on account of (7). Therefore there must be a $\xi < \nu$ such that $\mathfrak{G} \subseteq \mathfrak{H}_\xi$, and hence, by (6), $\mathfrak{c}(\mathfrak{G}) < \mathfrak{n}$.

Now assume that (16) has been proved for a given positive integer l, and consider a family $\mathfrak{G} \subseteq \mathfrak{F}$ of mutually exclusive sets of the $(l + 1)$st order. Let X be any set of $\mathfrak{G}$, and let $\mathfrak{H}_{\xi_1}, \mathfrak{H}_{\xi_2}, \cdots, \mathfrak{H}_{\xi_{l+1}}$ $(\xi_1 < \xi_2 < \cdots < \xi_{l+1} < \nu)$ be those families $\mathfrak{H}_\xi$ which by (13) contain among their elements a factor of X. By (14) every set of $\mathfrak{G}$ must have a factor belonging to at least one of the families $\mathfrak{H}_{\xi_1}, \cdots, \mathfrak{H}_{\xi_{l+1}}$, and thus also to their union

$$\mathfrak{H}^* = \mathfrak{H}_{\xi_1} \cup \cdots \cup \mathfrak{H}_{\xi_{l+1}} . \tag{17}$$

For every set Z of H^* let us denote by $\mathfrak{G}(Z)$ the family of all those sets $Y \in \mathfrak{G}$ which have Z as a factor. We have thus a decomposition of $\mathfrak{G}$ in subfamilies $\mathfrak{G}(Z)$ (which are not unnecessarily mutually exclusive):

$$\mathfrak{G} = \bigcup_{Z \in \mathfrak{H}^*} \mathfrak{G}(Z).$$

Hence

$$\mathfrak{c}(\mathfrak{G}) \leqq \sum_{Z \in \mathfrak{H}^*} \mathfrak{c}[\mathfrak{G}(Z)]. \tag{18}$$

Consider a particular family $\mathfrak{G}(Z)$ where Z is any set of $\mathfrak{H}^*$. If Y is a set of $\mathfrak{G}(Z)$, it has Z as a factor. Denote by Y^* the intersection of the remaining factors of Y, and by $\mathfrak{G}^*(Z)$ the family of all sets Z^* thus obtained. $\mathfrak{G}(Z)$ being a family of mutually exclusive sets, we easily infer from (13) and (15) that the family $\mathfrak{G}^*(Z)$ is also a family of mutually exclusive sets; furthermore that the correspondence $Y \dashrightarrow Y^*$ between the sets of $\mathfrak{G}(Z)$ and $\mathfrak{G}^*(Z)$ is one-to-one, and that therefore these two families have the same power. On the

other hand, $\mathfrak{G}^*(Z)$ consists of sets of the l^{th} order; thus, by applying to $\mathfrak{G}^*(Z)$ our inductive premise, we obtain:

(19) $$\mathfrak{c}[\mathfrak{G}(Z)] = \mathfrak{c}[\mathfrak{G}^*(Z)] < n \quad \text{for every} \quad Z \in \mathfrak{H}^*.$$

$\mathfrak{n}$ being a regular, and thus an infinite, cardinal number, we also have, by (6) and (17).

$$\mathfrak{c}(\mathfrak{H}^*) \leqq \mathfrak{c}(\mathfrak{H}_{\xi_1}) + \cdots + \mathfrak{c}(\mathfrak{H}_{\xi_{l+1}}) < \mathfrak{n};$$

and hence, in view of (19),

(20) $$\sum_{Z \in \mathfrak{H}^*} \mathfrak{c}[\mathfrak{G}(Z)] < \mathfrak{n}.$$

From (18) and (20) it follows at once that $\mathfrak{G}$ has power $< \mathfrak{n}$. Thus (16) holds for every positive integer l.

We can now extend (16) in the following way:

(21) Every family $\mathfrak{G} \subseteq \mathfrak{F}$ of mutually exclusive sets of any finite orders has power $< \mathfrak{n}$.

In fact, denote by $\mathfrak{G}_l$ the family of those sets $X \in \mathfrak{G}$ which are of the l^{th} order. We obtain the decomposition

$$\mathfrak{G} = \mathfrak{G}_1 \cup \cdots \cup \mathfrak{G}_l \cup \cdots,$$

whence

$$\mathfrak{c}(\mathfrak{G}) \leqq \mathfrak{c}(\mathfrak{G}_1) + \cdots + \mathfrak{c}(\mathfrak{G}_l) + \cdots.$$

(The families $\mathfrak{G}_1, \cdots, \mathfrak{G}_l, \cdots$ are not necessarily mutually exclusive.) On the other hand, we have by (16):

$$C(\mathfrak{G}_l) < \mathfrak{n} \text{ for every positive integer } l.$$

Hence, $\mathfrak{n}$ being by hypothesis a regular number $> \aleph_0$, we easily obtain:

$$\mathfrak{c}(\mathfrak{G}) \leqq \mathfrak{c}(\mathfrak{G}_1) + \cdots + \mathfrak{c}(\mathfrak{G}_l) + \cdots < \mathfrak{n}.$$

Finally we can show that

(22) Every family $\mathfrak{G} \subseteq \mathfrak{F}$ of mutually exclusive sets has 0 power $< \mathfrak{n}$.

For, by (11), every set $Y \in \mathfrak{G}$ is a union of finite intersections of sets $X \in \mathfrak{H}$ and their complements $N - X$. Furthermore, from (5) and (10) it follows that the complement $N - X$ of a set $X \in \mathfrak{H}$ is a union (not necessarily finite) of sets of $\mathfrak{H}$. Therefore the set Y can be represented as a union of finite intersections of sets $X \in \mathfrak{H}$. Hence we can correlate with every set $Y \in \mathfrak{G}$ (which is not empty) a non-empty subset $Y^* \subseteq Y$ of finite order. The family $\mathfrak{G}^*$ of all the sets Y^* thus obtained has clearly the same power as $\mathfrak{G}$, and by (21) this power is $< \mathfrak{n}$.

From the definition of $\mathfrak{d}$, (22) implies:

$$\mathfrak{d}(\mathfrak{G}) \leqq \mathfrak{n};$$

and this formula together with (12) gives:

(23) $$\mathfrak{c}(\mathfrak{F}) = \mathfrak{d}(\mathfrak{F}) = n.$$

Thus our theorem has been proved for a particular set N of power $\mathfrak{n}$. Now consider an arbitrary set S of power $\geqq \mathfrak{n}$. The set S contains a subset N_1 of power $\mathfrak{n}$. We can establish a one-to-one correspondence between the subsets of N and those of N_1, and by means of this correspondence we can construct a field $\mathfrak{F}_1$ of sets $X \subseteq N_1 \subseteq S$ which satisfies (23). This brings the proof to an end.

COROLLARY 1. *If $\mathfrak{n}$ is a regular number $>\aleph_0$, then for every number $\aleph \geqq \mathfrak{n}$ there exists a topological space S of power $\aleph$ such that the family $\mathfrak{G}$ of all open sets (or of all sets which are both open and closed, or of all regular open sets) satisfies the condition:* $\mathfrak{d}(G) = \mathfrak{n}$.

PROOF. Consider any set S of power $\aleph$. By Theorem 2 there is a field $\mathfrak{F}$ of subsets of S such that $\mathfrak{d}(\mathfrak{F}) = \mathfrak{n}$ and that every family $\mathfrak{H} \subseteq \mathfrak{F}$ of mutually exclusive sets has power $<\mathfrak{n}$. We can assume that $S \in \mathfrak{F}$; for otherwise, we could replace $\mathfrak{F}$ by the field $\mathfrak{F}'$ consisting of all the sets $X \in \mathfrak{F}$ and their complements $S - X$, and we could easily show that $\mathfrak{F}'$ still satisfies the conclusion of Theorem 2. Now we correlate with every subset $X \in S$ a new subset $\bar{X} \in S$, the closure of X, by defining $\bar{X}$ as the intersection of all the sets $Y \in \mathfrak{F}$ which contain $X(Y \supseteqq X)$. It is easily seen that with this definition S becomes a topological space.

In this space, F is contained in the family $\mathfrak{G}'$ of all those sets which are both open and closed, and the family G of all open sets is constituted by all the unions of the sets $X \in \mathfrak{F}$. Hence it follows that $\mathfrak{d}(G) = \mathfrak{d}(F) = \mathfrak{n}$, and that $\mathfrak{G}$, like $\mathfrak{F}$, does not contain any family $\mathfrak{H}$ of mutually exclusive sets with power $\mathfrak{c}(\mathfrak{H}) = \mathfrak{n}$. Finally it may be noticed that $\mathfrak{G}'$ and the family $\mathfrak{G}''$ of all regular open sets also have these two properties, for we clearly have

$$F \subset \mathfrak{G}' \subset \mathfrak{G}'' \subset \mathfrak{G}.$$

COROLLARY 2. *If $\mathfrak{n}$ is a regular number $>\aleph_0$, then there exists a complete Boolean algebra B such that* $\mathfrak{d}(B) = \mathfrak{n}$.

PROOF. This corollary follows directly from Corollary 1 since, as is well known, the regular open sets of an arbitrary topological space form a complete Boolean algebra,[3] and any two elements of this algebra are disjoint if, and only if, they are disjoint in the usual set-theoretic sense.

Corollary 2 formulates a condition which is necessary for a cardinal number $\mathfrak{n}$ to be regular and $>\aleph_0$. From Theorem 1 it follows that this condition is at the same time a sufficient one. It is easily seen that in this necessary and sufficient condition the term "Complete Boolean algebra" can be replaced by "partially ordered set", "ring (or field) of sets", "family of all open sets of topological space," and so on. If we restrict ourselves to the case of limit numbers, we obtain a necessary and sufficient condition for a number $\mathfrak{n} > \aleph_0$ to be weekly inaccessible.

---

[3] This result was first stated in A. Tarski, Les fondements de la géometrie de corps, Commemoration of the first Polish Math. Congress, Kracow 1929, p. 42; see also G. Birkhoff op. cit.. p. 102.

Finally we give a result of a more special nature:

COROLLARY 3. *The following two sentences are equivalent:*

(*i*) *In every topological space of power* $\leqq 2^{\aleph_0}$ *there exists a family of mutually exclusive open sets with a maximal power.*

(*ii*) *There is no weakly inaccessible number* $\mathfrak{n}$ *which is* $>\aleph_0$ *and* $\leqq 2^{\aleph_0}$.

PROOF. From Corollary 1 it follows immediately that (i) implies (ii); the implication in the opposite direction can be easily derived from Theorem 1.

GENERAL REMARKS ON INACCESSIBLE NUMBERS

In connection with the last corollary it should be noticed that the problem as to whether there exist weakly inaccessible numbers i.e. regular limit numbers which are $>\mathfrak{n}$ and $<2^{\mathfrak{n}}$ for an infinite number $\mathfrak{n}$ is so far unsolved and probably can not be solved at all within the present systems of general set theory. By definition the weakly inaccessible numbers can not be obtained from smaller ones by such operations as those of infinite addition or of passage from one number to the next greater number. However it is by no means settled that they can not be obtained from smaller ones by means of the other arithmetical operations, namely multiplication and exponentiation. For this reason we single out among the weakly inaccessible numbers a more special class the so called strongly inaccessible numbers, i.e., the numbers which can not be obtained from smaller ones by an arithmetical operation. While e.g. $2^{\aleph_0}$ is clearly not a strongly inaccessible number, it is not known whether this number is weakly inaccessible.

If we enrich the axiom system of set theory by adding the so-called generalized hypothesis of Cantor (which asserts that there is no cardinal number $>\mathfrak{n}$ and $<2^{\mathfrak{n}}$ for any infinite number $\mathfrak{n}$), we can easily show that the two kinds of inaccessible numbers coincide. However nothing compels us to regard the generalized hypothesis of Cantor as the only possible basis for set-theoretic investigations, and we can equally well consider the possibility of enriching the axioms of set theory by other axioms which contradict the hypothesis of Cantor. For instance it seems quite plausible that the following hypothesis would constitute a consistent and fertile addition to the set theoretical axioms:

Hypothesis of inaccessible numbers: For every infinite number n, 2^{n} is the smallest weakly inaccessible number $>\mathfrak{n}$.

Furthermore we should like to point out that many set theoretical problems are known at present whose solution involves the notion of an inaccessible number. The first problems of this kind were formulated more than thirty years ago; their number has however considerably increased in recent years. Like the problem solved in the present paper most of these problems can be presented in the following form: Is it true that a certain cardinal number $\mathfrak{n}$ has property P?

We want to give here a few examples of such problems:

PROBLEM 1. (*The representation problem.*) *Is it true that every* $\mathfrak{n}$*-additive and* $\mathfrak{n}$*-distributative Boolean algebra is isomorphic with an* $\mathfrak{n}$*-additive field of sets?* (A Boolean algebra B is called $\mathfrak{n}$-additive if for every set $X \subseteq B$ with $\mathfrak{c}(X) < \mathfrak{n}$

there is an element $y \in B$ such that $y = \bigcup_{x \in X} x$. An $\mathfrak{n}$-additive Boolean algebra is called $\mathfrak{n}$-distributive if

$$\bigcap_{i \in \mathfrak{H}} \bigcup_{j \in \mathfrak{G}_i} x_{i,j} = \bigcup_{f} \bigcap_{i \in \mathfrak{H}} x_{i,f(i)} ,$$

where $\mathfrak{H}$ is any non-empty set with $\mathfrak{c}(\mathfrak{H}) < \mathfrak{n}$; $\mathfrak{G}_i$ (for $i \in \mathfrak{H}$) are any non-empty sets with $\mathfrak{c}(\mathfrak{G}_i) < \mathfrak{n}$; $x_{i,j}$ is always an element of B, and f runs through all functions which correlate with every element $i \in \mathfrak{H}$ an element $j_i \in \mathfrak{G}_i$. The number of functions f is in general $\geqq \mathfrak{n}$, but it is assumed that the existence of $\bigcap_{i \in H} \bigcup_{j \in j_i} X_{i,j}$ implies that of $\bigcup_f \bigcap_{i \in H} x_{i,f(i)}$.)

PROBLEM 2. (*The prime ideal problem.*) *Is it true that the field of all subsets of a set N with power $\mathfrak{c}(N) = \mathfrak{n}$ contains an $\mathfrak{n}$-additive ideal which is not a principal ideal?* (A family $\mathfrak{F}$ is called n-additive if for every family $\mathfrak{G} \subseteq \mathfrak{F}$ with $\mathfrak{c}(\mathfrak{G}) < \mathfrak{n}$ the union $\bigcup_{X \in \mathfrak{G}} X$ also belongs to $\mathfrak{F}$.)

This problem can also be formulated as that of the existence of an $\mathfrak{n}$-additive non trivial two-valued measure defined over all the subsets of a set N with $\mathfrak{c}(N) = \mathfrak{n}$.

PROBLEM 3. (*The set-function problem.*) *Is it true that there exists an $\mathfrak{n}$-additive and $\mathfrak{n}$-multiplicative set-function defined over all sybsets of a set N of power $\mathfrak{n}$, which is not absolutely additive and absolutely multiplicative?* (By a set function we mean here a function G which correlates with every set X of a certain family $\mathfrak{F}$ another set $G(X)$ which need not belong to the same family. A set function G is called $\mathfrak{n}$-additive or $\mathfrak{n}$-multiplicative, if for every family $\mathfrak{H} \subseteq \mathfrak{F}$ with $\mathfrak{c}(\mathfrak{H}) < \mathfrak{n}$ we have:

$$G(\bigcup_{X \in \mathfrak{H}} X) = \bigcup_{X \in \mathfrak{H}} (G(X)) \quad \text{or} \quad G(\bigcap_{X \in \mathfrak{H}} X) = \bigcap_{X \in \mathfrak{H}} (G(X)),$$

respectively. If these formulas hold for every family $\mathfrak{H} \leqq \mathfrak{F}$, the set function is called absolutely additive or absolutely multiplicative.)

PROBLEM 4. (*The graph problem.*) *Is it true that if a complete graph G of power $\mathfrak{n}$ is split into two graphs G_1 and G_2, at least one of them contains a subgraph of power $\mathfrak{n}$?* (A graph is to be defined as an arbitrary set of non-ordered couples (x, y) with $x \neq y$. By a complete graph of power $\mathfrak{n}$ we mean the set of all such couples formed from the elements of a set N of power $\mathfrak{n}$.)

PROBLEM 5. (*The ordering problem.*) *Is it true that every ordered set N of power $\mathfrak{n}$ contains a subset X of power $\mathfrak{n}$, which is either well ordered, or becomes well ordered if we invert the ordering relation.*

PROBLEM 6. (*Ramification problem.*) *Let ν be the smallest ordinal number such that the power of all ordinals $\xi < \nu$ is $\mathfrak{n}$. Is it true that every ramification system of the ν^{th} order, in which the set of all elements of the ξ^{th} order has power $< \mathfrak{n}$ for every $\xi < \nu$, contains a well-ordered subset of the type ν.* (By a ramification system S we understand a partially ordered set which has the property that, for every $x \in S$, the set $S(x)$ of all elements $y \leqq x$ is well ordered; If the set $S(x)$ is of the type ξ the element x is said to be of the ξ^{th} order. The order of the

whole set S is the smallest ordinal number greater than the order of all elements of S.)

None of these six problems has yet been entirely solved. It can be shown that the solution of these problems is positive for $\mathfrak{n} = \aleph_0$ and is negative for every infinite number $\mathfrak{n} > \aleph_0$ which is not strongly inaccessible, in the case of problems 1–5. (In the case of problem 6 it has been only shown that the solution is negative if $\mathfrak{n}$ is not inaccessible and the generalized Cantor hypothesis holds.) All the problems remain open in case of strongly inaccessible numbers $> \aleph_0$.[4]

This situation is rather typical of the problems involving the notion of an inaccsssible number, which we have here in mind. Most of them so far have resisted all attempts at solution in the case in which $\mathfrak{n}$ is an inaccessible number $> \aleph_0$; it depends, however, on the nature of the problem whether strongly or weakly inaccessible numbers are involved. Ths situation differs slightly in connection with certain problems from the theory of ordered sets. Here the solution is positive for $\aleph_0$, and for all regular numbers which are not weakly inaccessible, and is negative for infinite singular numbers; but the problem again remains open in the case of weakly inaccessible numbers $> \aleph_0$.

The difficulties which we meet in attempting to solve the problems under consideration do not seem to depend essentially on the nature of inaccessible

[4] The solution of Problem 1 was given for $\mathfrak{n} = \aleph_0$ by M. H. Stone (see G. Birkhoff op. cit. p. 89.) The solution for numbers which are not inaccessible and $> \aleph_0$ was recently found by A. Tarski and has not yet been published.

For the solution of Problem 2. see A. Tarski, Fund. Math. Vol. 15, p. 42–50. (Une contribution à la théorie de la mesure) (the case $\mathfrak{n} = \aleph_0$). For the case when $\mathfrak{n}$ is not inaccessible, see A. Tarski, Fund. Math. Vol. 30 (1938) p. 150 (Dritter überderkungnatz.)

The solution of Problem 3 for $\mathfrak{n} = \aleph_0$ was given by S. Ulam, Fund. Math. Vol. 16 p. 140–150. (Zur Masstheorie in der allgemeinen Mengenlehre.) The solution for numbers which are not inaccessible follows from a general theorem of A. Tarski; C. R. Soc. Varsovie, Vol. 30 p. 158 (Theorem 2.18).

The solution of Problem 4 was given for $\mathfrak{n} = \aleph_0$ by Ramsay on a problem of formal logic, Proc. London Math. Soc. (2), 30; and for the numbers $\mathfrak{n} > \aleph_0$ which are not inaccessible by P. Erdös, appear in Revista de Tucuman).

The solution of Problem 5 is obvious for $\mathfrak{n} = \aleph_0$; for the numbers $\mathfrak{n} > \aleph_0$ which are not inaccessible the solution was given by Hausdorff, Mengenlehre (1914) p. 145–146. He does not state the solution explicitly, but it can be deduced easily from his results.

The solution of Problem 6 was given for $\mathfrak{n} = \aleph_0$ by D. König, Über eine Schlussweise aus dem endlichen ins unendliche, Acta Szeged, 3, p. 121–130. For the numbers $\mathfrak{n} > \aleph_0$ which are not inaccessible it was given by Aronsajn. It can be shown that the positive solution of Problem 1 for inaccessible numbers $> \aleph_0$ would imply the positive solution of Problem 2; the positive solution of Problem 2 implies that of Problems 3, 4, and 5; also the positive solution of 3 implies that of 2, so that 2 and 3 are equivalent. Further, the positive solution of 4 implies that of 5, and the positive solution of Problem 6 for strongly inaccessible numbers can be deduced from that of Problem 2 (however this solution can also be obtained from weaker hypotheses and can be extended to all inaccessible numbers). Also the positive solution of Problem 6 implies that of Problems 4 and 5. Finally the positive solution of Problem 6 implies the positive solution of Problem 1 in the special case when the Boolean algebra contains only $\mathfrak{n}$ elements. The proof of these equivalences is as yet unpublished.

numbers. In most cases the difficulties seem to arise from lack of devices which enable us to construct maximal sets which are closed under certain infinite operations. It is quite possible that a complete solution of these problems would require new axioms which would differ considerably in their character not only from the usual axioms of set theory, but also from those hypotheses whose inclusion among the axioms has previously been discussed in the literature and mentioned previously in this paper (e.g., the existential axioms which secure the existence of inaccessible numbers, or from hypotheses like that of Cantor which establish arithmetical relations between the cardinal numbers.)

If we now compare the problem which has been actually solved in this paper with those which we have recently discussed, we see that the peculiarity of our problem consists in two facts. First, our problem has been solved for all cardinal numbers, although the inaccessible numbers are essentially involved in the solution. And secondly the number $\aleph_0$ behaves in the discussion of the problem like a singular limit number, and in a directly opposite way to the other regular or inaccessible numbers.

UNIVERSITY OF PENNSYLVANIA AND INSTITUTE FOR ADVANCED STUDY

Annals of Mathematics
Vol. 44, No. 2, April, 1943

ON SOME CONVERGENCE PROPERTIES OF THE INTERPOLATION POLYNOMIALS

P. Erdös

(Received June 29, 1942)

It is well known that there exist continuous functions whose Lagrange interpolation polynomials taken at the roots of the Tchebycheff polynomials $T_n(x)$ diverge everywhere in $(-1, +1)$.[1] On the other hand a few years ago S. Bernstein proved the following result[2]: Let $f(x)$ be any continuous function; then to every $c > 0$ there exists a sequence of polynomials $\varphi_n(x)$ where $\varphi_n(x)$ is of degree $n - 1$ and it coincides with $f(x)$ at, at least $n - cn$ roots of $T_n(x)$ and $\varphi_n(x) \to f(x)$ uniformly in $(-1, +1)$.

Fejér proved the following theorem[3]: Let the fundamental points of the interpolation be a normal[4] point group

$$\begin{array}{c} x_1^{(1)} \\ x_1^{(2)},\quad x_2^{(2)} \\ \cdots\cdots\cdots\cdots; \end{array}$$

then for every continuous $f(x)$ there exists a sequence of polynomials $\varphi_n(x)$ of degree $\leqq 2n - 1$ such that $\varphi_n(x_i^{(n)}) = f(x_i^{(n)})$, $i = 1, 2, \cdots n$ and $\varphi_n(x) \to f(x)$ uniformly in $(-1, +1)$. In the present paper we are going to prove the following more general

THEOREM 1. *Let the point group be such that the fundamental functions $l_k^{(n)}(x)$ are uniformly bounded in $(-1, +1)$. Then to every continuous function $f(x)$ and $c > 0$ there exists a sequence of polynomials $\varphi_n(x)$, such that, 1) the degree of $\varphi_n(x)$ is $\leqq n(1 + c)$, 2) $\varphi_n(x_i^{(n)}) = f(x_i^{(n)})$, $i = 1, 2 \cdots n$, 3) $\varphi_n(x) \to f(x)$ uniformly in $(-1, +1)$.*

Theorem 1 generalizes the result of Fejér in two directions; first the point group is more general since it can be shown[5] that the fundamental functions are uniformly bounded for normal point groups, and secondly the degree of $\varphi_n(x)$ is lowered from $2n - 1$ to $n(1 + c)$.

Theorem 1 does not directly generalize the result of S. Bernstein, but we can prove the following

THEOREM 2. *Let the $x_i^{(n)}$ be such that the fundamental functions are uniformly bounded in $(-1, +1)$; then to every continuous function $f(x)$ there exists a sequence of polynomials $\varphi_n(x)$* of degree $\leqq n - 1$ *which coincides with $f(x)$ at, at least n-cn points $x_i^{(n)}$ and $\varphi_n(x) \to f(x)$.*

---

[1] G. Grünwald, Annals of Math. Vol. 37, (1936), p. 908–918.

[2] S. Bernstein, Comptes Rendus de l'Acad. des Sciences Vol.

[3] L. Fejér, Amer. Math. Monthly Vol. 41 p. 12.

[4] Ibid.

[5] Fejér proves this only for the so called strongly normal point groups (ibid). The proof for normal point groups is much more complicated and we do not give it here.

We are not going to give a proof of Theorem 2.

The following problem is due to Fejér: Let the $x_i^{(n)}$ be the equidistant abscissae that is $x_i^{(n)} = -1 + \frac{2i-1}{n}$, $i = 1, 2, \cdots n$. The question is, does there exist to every continuous $f(x)$ a sequence of polynomials $\varphi_n(x)$ of degree $< 2n$ such that $\varphi_n(x_i^{(n)}) = f(x_i^{(n)})$ and $\varphi_n(x) \to f(x)$. In other words, does his result proved for the normal point groups also hold for the equidistant point group. We prove the following

THEOREM 3. *To every continuous function $f(x)$ and to every c there exists a sequence of polynomials $\varphi_n(x)$ of degree $\leqq \frac{\pi}{2} n(1+c)$ such that $\varphi_n(x_i^{(n)}) = f(x_i^{(n)})$ and $\varphi_n(x) \to f(x)$ uniformly in $(-1, 1)$, and it can be shown that the constant $\frac{\pi}{2}$ is the best possible.*

Throughout this paper the c's denote absolute constants not necessarily the same. If there is no danger of confusion we will omit the upper index n in $x_i^{(n)}$, $l_k^{(n)}(x)$ etc.

To prove Theorem 1 we need two lemmas.

LEMMA 1. *Let the point group be such that the fundamental functions are uniformly bounded in $(-1, 1)$ and put $\cos \vartheta_i = x_i$, $x_1 < x_2 < \cdots < x_n$, $\vartheta_1 > \vartheta_2 > \cdots > \vartheta_n$; then*

$$\vartheta_i - \vartheta_{i+1} > \frac{c}{n}.\ [6]$$

PROOF. Let $|l_i(x)| < D$, $i = 1, 2, \cdots n$. By a well known theorem of S. Bernstein[7] $|d/d\vartheta\, l_i(\cos\vartheta)| \leqq nD$ and since $l_i(x_i) = 1$, $l_i(x_{i\pm1}) = 0$, we have finally

$$\vartheta_i - \vartheta_{i+1} \geqq \frac{1}{Dn}.$$

LEMMA 2. *Let $-1 \leqq y \leqq 1$, $\cos\theta = y$. Then there exists a polynomial $h_y^{(m)}(x)$ of degree $\leqq 2m$ such that $h_y^{(m)}(y) = 1$, $|h_y(x)| \leqq c$, $-1 \leqq x \leqq 1$ and for $\theta - \theta_0 > \frac{A}{m}$*

$$|h_y^{(m)}(\cos\theta_0)| < c_1 \min\left(1, \frac{1}{m^2(\theta-\theta_0)^2}\right).$$

Denote by $X_i^{(m)}$ and X_{i+1}^{m} the roots of $T_m(x)$ for which $X_i^{(m)} \leqq y \leqq X_{i+1}^{(m)}$. It is easy to see that

---

[6] If the fundamental functions are uniformly bounded we have

$$\frac{c_1}{n} < \vartheta_i - \vartheta_{i+1} < \frac{c_2}{n}.$$

But the upper estimate is not needed here. (Erdös-Turán, Annals of Math. Vol. 39 (1940) p. 706–707.)

[7] S. Bernstein, Belg. Mém. 1912 p. 19.

$$L_i^{(m)}(y) + L_{i+1}^{(m)}(y) \geqq 1,$$ [8]

where $L_i^{(m)}(y)$ denotes the fundamental polynomials belonging to the roots of $T_m(x)$. Without loss of generality we may assume $L_i(y) \geqq \frac{1}{2}$. It is well known that $| L_i^{(m)}(x) | \leqq \sqrt{2}, -1 \leqq x \leqq 1$.[9] Thus since $\theta_i - \theta_{i+1} = \frac{\pi}{m}$ ($\cos \theta_i = X_i$) our lemma will be proved if we can show that for $| \theta_i - \theta_0 | > \frac{A}{m}$

$$| h_y^{(m)}(\cos \theta_0) | = | L_i^{(m)}(x_0)/L_i^{(m)}(y) |^2 < \frac{c}{A^2}.$$

But

$$| h_y^{(m)}(\cos \theta_0) | \leqq \frac{4}{m^2(\theta_i - \theta_0)^2} < \frac{c}{A^2}.$$

PROOF of Theorem 1. Let $\psi_{n-1}(x)$ be a polynomial of degree n-1 such that

$$| f(x) - \psi_{n-1}(x) | < \epsilon, \ -1 \leqq x \leqq 1.$$

Put $f(x_i) - \psi_{n-1}(x_i) = \epsilon_i$. Consider the polynomial of degree $\leqq n(1 + c)$ such that

$$\varphi_{n-1}(x) = \psi_{n-1}(x) + \sum_{i=1}^{n} \epsilon_i l_i(x) h_{x_i}^{(m)}(x), \qquad m = \left[\frac{cn}{2}\right].$$

Clearly $\varphi_{n-1}(x_i) = f(x_i)$, $i = 1, 2 \cdots n$. We shall prove that $\varphi_{n-1}(x) \rightarrow f(x)$ uniformly in $(-1, 1)$. It suffices to show that

$$| g(x) | = \left| \sum_{i=1}^{n} \epsilon_i l_i(x) h_{x_i}^{(m)}(x) \right| < c\epsilon, \qquad -1 \leqq x \leqq 1.$$

Now

$$| g(x) | < c\epsilon \sum_{i=1}^{n} | h_{x_i}^{(m)}(x) | = c\epsilon \sum_{x_i \geqq x} | h_{x_i}^{(m)}(x) |$$
$$+ c\epsilon \sum_{x_i \geqq x} | h_{x_i}^{(m)}(x) | = c\epsilon \left(\sum\nolimits_1 + \sum\nolimits_2\right)$$

Thus we only have to show that $\sum_1 + \sum_2 < c_1$. By Lemma 1,

$$\sum\nolimits_1 < \sum_r | h_{x+k_r}^{m}(x) |,$$

where $| \text{arc} \cos(x + k_r) - \text{arc} \cos x | > (rc)/n$. Thus by Lemma 2

$$\sum\nolimits_1 < \sum_r \frac{c_3}{r^2 c^2} < c_4.$$

Similarly we obtain $\sum_2 < c_2$, which completes the proof of Theorem 1

[8] Erdös-Turán, Annals of Math. Vol. 41, (1941) p. 529, Lemma IV.

[9] L. Fejér, Mathematische Annalen, Vol. 106, (1932) p. 5.

Theorem 1 does not give a necessary and sufficient condition for the existence of a sequence of polynomials $\varphi_n(x)$ of degree $\leqq n(1 + c)$ with $\varphi_n(x_i) = f(x_i)$ and $\varphi_n(x) \to f(x)$ uniformly in $(-1, 1)$. To obtain such a condition let $x_i^{(n)}$ be a point group, put $\cos \vartheta_i^{(n)} = x_i^{(n)}$ and denote by $N_n\ (a, b)$ the number of the ϑ_i in (a, b), We have the following:

THEOREM 4. *A necessary and sufficient condition that to every continuous function $f(x)$ and to every $c > 0$ there exists a sequence of polynomials $\varphi_n(x)$ of degree $\leqq n(1 + c)$ such that $\varphi_n(x_i) = f(x_i)$ and $\varphi_n(x) \to f(x)$ uniformly in $(-1, 1)$ is that if $n\ (b_n - a_n) \to \infty$, $0 \leqq a_n < b_n \leqq \pi$*

$$(1)\ \limsup. \frac{N_n(a_n, b_n)}{n(b_n - a_n)} \leqslant \frac{1}{\pi} \text{ and } \liminf. (\vartheta_i - \vartheta_{i+1})n > 0,\ (n \to \infty\ i \text{ arbitrary})$$

Condition (1) *states that the number of ϑ_i in (a_n, b_n) can not be much greater than the number of roots of $T_n(x)$ in (a_n, b_n). If the fundamental functions $l_k(x)$ are uniformly bounded* (1) *is satisfied, for then we have*

$$\lim \frac{N_n(a_n, b_n)}{n(b_n - a_n)} = \frac{1}{\pi}, \qquad n(b_n - a_n) \to \infty \text{ [10]}$$

We do not give the proof of Theorem 4, but the following proof of Theorem 3 can by a simple modification be applied to it.

PROOF of Theorem 3. Here the fundamental points are

$$x_i^{(n)} = -1 + \frac{2i - 1}{n}$$

First we prove the existence for every n and $c > 0$ of $m = \frac{\pi}{2} n(1 + c)$ points. $y_i^{(m)}$, $i = 1, 2, \cdots n$ such that (I) the $x_i^{(n)}$ occur among the $y_i^{(m)}$ (II) the fundamental functions $L_k(x)$, $k = 1, 2, \cdots m$ are uniformly bounded in $(-1, 1)$ (The $L_k(x)$ are the fundamental functions belonging to the $y_i^{(m)}$). Having constructed the $y_i^{(m)}$ satisfying (I) and (II) we immediately obtain Theorem 3 by applying Theorem 1.

To construct the $y_i^{(m)}$ we first remark that by putting

$$\cos \vartheta_i = -1 + \frac{2i - 1}{n}, \qquad i = 1, 2, \cdots n$$

we obtain by a simple calculation

$$\vartheta_i - \vartheta_{i+1} > \frac{\pi}{m}$$

Now we construct a sequence $y_i^{(m)}$, $i = 1, 2, \cdots m$ such that (1) the $x_i^{(n)}$ occur among the $y_i^{(m)}$ (2) put $\cos \theta_i = y_i$ then $\theta_i = \frac{2i - 1}{m} \frac{\pi}{2} + \frac{d_i}{m}$ where $\sum_{i=1}^{k} d_i$ is uniformly bounded (3) $\theta_i - \theta_{i+1} \geqq \frac{\pi}{4m}$. (2) and (3) insure that the y_i are "very nearly" the roots of $T_m(x)$.

[10] Erdös-Turán, ibid., p. 519.

We construct the y_i as follows: Suppose $y_i < y_2 < \cdots < y_{i-1}$ are already constructed. We further make the hypothesis that if ϑ_r ($\cos \vartheta_r = x_r$) is the greatest $\vartheta < \theta_{i-1}$ then $\theta_{i-1} - \vartheta_r > \frac{\pi}{4m}$. If $\sum_{j=1}^{i-1} d_j < 0$ we choose for y_i either the least $x_r > y_{i-1}$, or if $\vartheta_r < \theta_{i-1} - \frac{4\pi}{m}$ we put $\theta_i = \theta_{i-1} - \frac{2\pi}{m}$. Thus θ_i does not come nearer than $\frac{\pi}{4m}$ to the greatest $\vartheta < \theta_i$. If $\sum_{j=1}^{i-1} d_j > 0$, $y_i = x_r$ if $\vartheta_r > \theta_{i-1} - \frac{\pi}{2m}$ and $\theta_{i-1} - \frac{\pi}{4m}$ otherwise. Thus in any case if ϑ_j is the greatest $\vartheta < \theta_i$ then $\theta_i - \vartheta_j > \frac{\pi}{4m}$. In this we can construct $y_1, y_2, \cdots y_m$. (1) and (3) are clearly satisfied and it is quite immediate that (2) is also satisfied. Now we have to show that the y_i's satisfying (1), (2), and (3) also satisfy (I) and (II). (I) is clearly satisfied, the proof that (II) is satisfied is slightly more difficult. Denote by $z_1, z_2, \cdots z_m$ the roots of $T_m(x)$ and by $L_k'(x)$ the fundamental functions belonging to the z_i. From (2) and (3) it follows by a simple calculation that

$$(2) \qquad c_1\omega'(y_k) < T_m'(z_k) < c_2\omega'(y_k), \qquad \omega(x) = \prod_{i=1}^{m}(x - y_i)$$

where c_1 and c_2 are independent of m and k. Denote

$$\max_{-1 \leq x \leq 1} |L_k(x)| = A_k, \qquad \max_{-1 \leq x \leq 1} L_k'(x) = B_k$$

Then again from (2) and (3) by a simple calculation [using (2)]

$$(3) \qquad c_3 > \frac{A_k}{B_k} < c_4 .$$

We know that[11]

$$(4) \qquad B_k < \sqrt{2}.$$

Thus from (3) and (4) we obtain (3), and this completes the proof of Theorem 3.

To obtain the second part of Theorem 3 we first have to prove

LEMMA 3. *Let $m = [(\pi/2)n(1 - \epsilon)]$, $\epsilon > 0$ fixed, independent of m and n, n odd. Let $\varphi_m(x)$ be a polynomial of degree m such that $\varphi_m(0) = 1$ and $\varphi_m(-1 + ((2i - 1)/n)) = 0$, $i = 1, 2, \cdots [(n - 1)/2], [(n + 3)/2] \cdots n$. Then*

$$\max_{-1 \leq x \leq 1} |\varphi_m(x)| > c_1^n, \qquad c_1 = c_1(\epsilon) > 1.$$

PROOF. We use the following lemma due to M. Riesz[12]: Let $\varphi_m(x)$ be a poly-

[11] L. Fejér, see footnote 9.

[12] M. Riesz, Jahresbericht der Deutschen Math. Vereinigung, (1915), p. 354–368.

nomial of degree m, it assumes its absolute maximum in $(-1, 1)$ at the point $x_0 = \cos \vartheta_0$. Let $x_i = \cos \vartheta_i$ be the nearest root of $\varphi_m(x)$ in $(-1, +1)$ then

$$|\vartheta_i - \vartheta_0| \geqq \frac{\pi}{2m}.$$

It immediately follows from this lemma that if x_i and x_{i+1} are the nearest roots including x_0, then we have

$$\vartheta_i - \vartheta_{i+1} \geqq \frac{\pi}{m}.$$

Put now $\cos \vartheta_i = -1 + (2i - 1)/n$. A simple calculation shows that there exists a constant $c_2 = c_2(\epsilon)$ such that if $-c_2 \leqq x_i < x_{i+1} \leqq c_2$ then

$$\vartheta_i - \vartheta_{i+1} < \frac{\left(1 - \frac{\epsilon}{2}\right)\pi}{m}.$$

Hence $\varphi_m(x)$ can not assume its absolute maximum for $-c_2 \leqq x \leqq c_2$ except if

$$x_{\frac{n-1}{2}} < x < x_{\frac{n+3}{2}} \qquad \text{(i. e. in the neighborhood of 0)}$$

Consider now a polynomial $h_m(x)$ with highest coefficient the same as that of $\varphi_m(x)$ whose roots are defined as follows: Let $-c_2 < z_i < c_2$ then $z_i = (1 + \delta)x_i$ where δ is chosen so small that

$$\theta_i - \theta_{i+1} < \frac{1 - \frac{\epsilon}{4}}{m}\pi \qquad (\cos \theta_i = z_i)$$

The other roots of $h_m(x)$ coincide with those of $\varphi_m(x)$. Clearly the degree of $h_m(x)$ is m. Define

$$g(x) = \left(x + \frac{1}{4m}\right)\left(x - \frac{1}{4m}\right)h_m(x).$$

By the lemma of M. Riesz $g(x)$ does not assume its absolute maximum in $(-c_2, c_2)$. It follows from the inequality of the arithmetic and geometric means that

(5) $$|g(x)| < |\varphi_m(x)| \text{ for } c_2(1 + \delta) \leqq |x| \leqq 1$$

Denote by $A(c_2)$ the number of x_i in $(-c_2, +c_2)$. We evidently have $A(c_2) > c_3 n$. Thus

(6) $$|g(0)| > |\varphi_m(0)|(1 + \delta)^{c_3 n} \frac{1}{16m^2} > \varphi_m(0)c_1^n = c_1^n (c_1 > 1)$$

But since $g(x)$ assumes its absolute maximum in $(-1, 1)$ for some $|x_0| > c_2(1 + \delta)$ we have by (5) and (6)

$$|\varphi_m(x_0)| > |g(x_0)| > c_1^n \quad \text{q.e.d.}$$

Let now $n_1, n_2, \cdots$ be an infinite sequence of odd integers, which tend to infinity sufficiently quickly. We define a polynomial $\psi_i(x)$ as follows

$$\psi_i\left(-1 + \frac{2j-1}{n_r}\right) = 0, \qquad r \leqq i, \qquad j \neq \frac{1+n_r}{2}, \qquad j \leqq n_r,$$

$$\psi_i(0) = 1, \qquad |\psi_i(x)| \leqq 2, \qquad -1 \leqq x \leqq 1.$$

From the approximation theorem of Weierstrass it follows that such a $\psi_i(x)$ exists. Consider now the continuous function

$$f(x) = \sum_{k=1}^{\infty} \frac{\psi_k(x)}{2^k}.$$

If the second part of Theorem 3 would not be true, we could find a sequence of polynomials $\varphi_i(x)$ of degree $\leqq n_i(\pi/2)(1-\epsilon)$ such that $\varphi_i(-1 + ((2j-1)/n_i)) = f[1 + ((2j-1)/n_i)]$ and $\varphi_i(x) \to f(x)$ uniformly in $(-1, 1)$. For $k > i$

$$\psi_k\left(-1 + \frac{2j-1}{n_i}\right) = 0, \qquad j \neq \frac{1+n_i}{2}.$$

Thus $\varphi_i(x)$ coincides with

$$\sum_{r=1}^{i-1} \frac{\psi_r(x)}{2^r} = g(x)$$

at the points $-1 + ((2j-1)/n_i)$, $j \neq ((1+n_i)/2)$.

Let now n_i tend to infinity so quickly that n_i is greater than the degree of $g(x)$. Then $\varphi_i(x)$ can be written as

$$\varphi_i(x) = \varphi_i^{(1)}(x) + \varphi_i^{(2)}(x),$$

where $\varphi_i^{(1)} = g(x)$, and $\varphi_i^{(2)}(x)$ is of degree $\leqq ((\pi/2) - c_1)n_2$ and $\varphi_i''(-1 + ((2j-1)/n_i)) = 0$, $j \neq ((1+n_i)/2)$, $j \leqq n_i$, also $\varphi_i^{(2)}(0) = \sum_{k \geqq i}((\psi_k(0))/2^k) = (1/2^{i-1})$. Thus by lemma 3

$$\max_{-1 \leqq x \leqq 1} |\varphi_i(x)| \geqq \max_{-1 \leqq x \leqq 1} |\varphi_i^{(2)}(x)| - 2 > \frac{c_2^{n_i}}{2^{i-1}} > c_3^{n_i} \text{ (c_2 and c_3 are} > 1$$

if n_i tends to infinity sufficiently quickly. Hence $\varphi_i(x)$ can not converge uniformly to $f(x)$, and this completes the proof of Theorem 3.

By a more complicated argument we could prove that a point x_0 exists such that $\varphi_n(x_0)$ diverges. We give only the sketch of the proof. Since $\max_{-1 \leqq x \leqq 1} |\varphi_{n_i}(x)| > (1+\delta)^{n_i}$ it follows from a theorem of Remes[13] that there exists in $(-1, 1)$ a set of measure $> c = c(\delta)$ such that on this set $|\varphi_{n_i}(x)| > (1 + (\delta/2))^{n_i}$. Then, it follows easily that there exist a point x_0 with $\limsup |\varphi_{n_i}(x_0)| = \infty$.

---

[13] E. Remès, *Sur une propriété extremale des polynomes de Tchebycheff.* Comm. de l'Institut des Sciences etc. Kharkov, (1936) série 4, XIII fasc. 1, p. 93–95.

By the same method we can prove the following:

THEOREM 5. *Let* $x_1^{(1)}$, $x_1^{(2)}, x_2^{(2)}$, $\dots$ *be a point group and put* $\cos(\vartheta_i^{(n)}) = x_i^{(n)}$. *Suppose that*

$$\liminf n(\vartheta_i^{(n)} - \vartheta_{i+1}^{(n)}) = \frac{\pi}{d}, \quad (n \to \infty, \ i \text{ arbitrary})$$

Then to every continuous $f(x)$ *and constant* $c > 0$ *there exists a sequence of polynomials* $\varphi_n(x)$ *of degree* $< d(1 + c)n$ *such that* $\varphi_n(x_i^{(n)}) = f(x_i^{(n)})$ *and* $\varphi_n(x) \to f(x)$ *uniformly in* $(-1, 1)$.

The constant d_1 of Theorem 5 is not best possible. We can obtain the best possible constant d_1 as follows: Let a_n and b_n be two arbitrary sequences of real numbers, such that $0 \leqq a_n < b_n \leqq \pi$, $n(b_n - a_n) \to \infty$. Then if $d < \infty$

$$\limsup \frac{N_n(a_n, b_n)}{n(b_n - a_n)} = \pi d_1 .$$

Lemma 3 would not suffice for the proof of Theorem 5. Here we need

LEMMA 4. *Let* $\varphi_n(x)$ *be a polynomial of degree* n, $\varphi_n(0) = 1$. *Let* $\psi(n)$ *be any function of* n *tending to infinity together with* n *and let* c_1 *be a constant independent of* n. *Then if* $\varphi_n(x)$ *is such that for every* $c_1 < A < \psi(n)$ *the number of roots of* $\varphi_n(\cos\vartheta)$ *in* $((\pi/2) - (A/n), (\pi/2) + (A/n))$ *is greater than* $[((1 + c_2)2)/\pi]$ *we have* $\max_{-1 \leqq x \leqq 1} |\varphi_n(x)| \to \infty$. *Our condition means that the number of roots of* $\varphi_n(x)$ *in the neighborhood of* O *is substantially larger than the number of roots of* $T_n(x)$. *The proof of Lemma 4 is similar, but more complicated than the proof of Lemma 3.*

UNIVERSITY OF PENNSYLVANIA

Annals of Mathematics
Vol. 44, No. 3, July, 1943

THE VARIATIONAL THEORY IN THE LARGE INCLUDING THE NON-REGULAR CASE—FIRST PAPER

By George Ewing and Marston Morse

(Received July 8, 1942)

Introduction

The classical minimum theory was first developed in the so-called regular case in which the Euler equations are non-singular. In obtaining the properties of the minimizing curve use was made of elementary extremals. Carathéodory studied a non-regular problem in 1906, but Tonelli was the first to make an extensive study of non-regular problems. This extension was real inasmuch as many important integrals fail to come under the regular case. Notable among such examples is the Jacobi least action integral I in the restricted problem of three bodies. A desire to understand I was one of the forces motivating this study. For it seems possible that an analysis of the contour manifolds of I in the Fréchet space of the admissible curves may reveal a hitherto unknown topological basis for the planetary orbits.

The theory in the large has a topological form independent of its application to functions of n variables to simple or double integrals, or to other problems. See Morse [1]. However, up to the present the application to the ordinary simple integral has presupposed the condition of positive regularity and has made definite use of broken extremals. It has been an open question whether the regularity condition could be relaxed in the theory in the large as it was by Tonelli in the minimum theory. The basic difficulties have now been met and the new theory includes the old.

A first change was in the theory of functions on a compact metric space. Two auxiliary metrics, a J-metric and an L-metric, had to be introduced to describe and establish the upper-reducibility and bounded compactness of J. Recall that a function $J(g)$ is "*boundedly compact*" if for each constant c the subset of points g, for which $J(g) \leqq c$, is compact. This change in the general theory has been made by Morse, [2].

The case of the simple integral without the regularity hypothesis is treated by Ewing and Morse in two papers of which this is the first. The second of these papers contains the major part of the new results. This first paper gives the conditions on the integrand f sufficient to insure the bounded compactness of J and the equivalence of convergence in length and J-length. The conditions of convexity and positive semi-normality which are used have received extensive attention. Besides Tonelli, a number of others including Hahn, Menger, Graves and McShane have made major contributions. We are particularly indebted to McShane as will be seen by our references. In this first paper we are thus concerned largely with exposition. It was necessary to bring the relevant material together in one place and put it in the form best suited to our purposes.

New proofs have been introduced, particularly in deriving the properties of the "figurative" from its convexity and the homogeneity of f, without using the differentiability of f. We also introduce the "pseudo limiting" curve and employ the concept of "regular convexity." A generalized Lindeberg theorem is stated with a reduction of hypotheses and a simplification of McShane's proof. The difficult problem of conditions sufficient for upper reducibility, and the proof of the generalized Euler theorem (the homotopy theorem) are left for the second paper, as well as the integration of the local hypotheses with the theory in the large.

1. The conditions on $f(x, r)$

The symbols x and r will designate sets $[x^{(1)}, \cdots, x^{(n)}]$ and $[r^{(1)}, \cdots, r^{(n)}]$ respectively. We shall refer to x and r as vectors and shall use the notation of vector analysis. Regarded as a point, x shall be restricted to a bounded closed connected set A in an n-dimensional cartesian space. Let R denote the cartesian space of points r. The integrand in our variational theory will be defined with the aid of a function $f(x, r)$ that is finite and single-valued for x on A and r on R. We shall condition f as follows:

I. *The function f shall be homogeneous in r in the sense that*

$$f(x, kr) = kf(x, r) \qquad k \geqq 0. \tag{1.1}$$

II. *The function f shall be convex in r for each x.*

A first consequence of II is that $f(x, r)$ is continuous in r. See Bonnesen und Fenchel [1], p. 19.

When f is convex in r at $x = c$, corresponding to c and to an arbitrary vector r there exists at least one vector $a(c, r_0)$ such that for arbitrary r and fixed r_0

$$f(c, r) \geqq a \cdot (r - r_0) + f(c, r_0) \tag{1.2}$$

where $a \cdot (r - r_0)$ is the scalar product of a by $(r - r_0)$.

We shall show that

$$a \cdot r_0 = f(c, r_0). \tag{1.3}$$

Upon setting $r = 2r_0$ and 0 respectively in (1.2), one finds that

$$f(c, r_0) \geqq a \cdot r_0 \qquad 0 \geqq f(c, r_0) - a \cdot r_0 .$$

Relation (1.3) follows.

We shall say that f is *positive semi-normal* at $x = c$ (written P.S.N.) if

$$f(c, r) > b \cdot r \qquad (r \neq 0)$$

for a vector constant b and any $r \neq 0$.

We begin with certain lemmas in which c is fixed.

LEMMA 1.1. *If $f(c, r) \geqq 0$ for all r, the subset in the space R on which $f(c, r) = 0$ is convex.*

Let r' and r'' be two points r at which $f(c, r) = 0$. On the line segment L

joining r' to r'', $f(c, r)$ is convex in r, and hence $f(c, r) \leqq 0$. But $f(c, r)$ is never negative by hypothesis, so that $f(c, r) \equiv 0$ on L, and the proof is complete.

We shall make use of the condition

$$f(c, r) + f(c, -r) > 0 \qquad (r \neq 0) \tag{1.4}$$

with c fixed and r an arbitrary non-null vector.

LEMMA 1.2. *If f is non-negative in r at $x = c$, and if (1.4) holds, there exists an $(n - 1)$ plane Π_{n-1} of the form $A \cdot r = 0$ such that $f(c, r) > 0$, where $A \cdot r \geqq 0$ and $r \neq 0$.*

When $n = 2$ the lemma follows readily. For $f(c, r)$ cannot vanish at diametrically opposite points r and $-r$ by virtue of the condition (1.4); and in accordance with Lemma 1.1 must vanish, if at all, in a convex domain bounded by two rays through the origin. The 1-plane Π_1 of the lemma is thus a suitably chosen straight line. To establish the lemma in general we assume that the lemma is true when $n = m - 1$, and show that it is true when $n = m$.

In the m-space of points r let Π_{m-1} be an arbitrary $(m - 1)$-plane through the origin. The function $f(c, r)$ will be a convex function ϕ of rectangular coordinates on Π_{m-1}, and the condition corresponding to (1.4) holds for ϕ, implying that ϕ does not vanish at diametrically opposite points of Π_{m-1}. By virtue of our inductive hypothesis there exists an $(m - 2)$-plane Π_{m-2} through the origin, with $\Pi_{m-2} \subset \Pi_{m-1}$ such that $f(c, r) > 0$ on Π_{m-1} on the closure of one side of Π_{m-2}, $r = 0$ excepted.

Let μ_{m-1} be an arbitrary half $(m - 1)$-plane with Π_{m-2} as a boundary. On no two diametrically opposite half planes of this type can $f(c, r) = 0$ on both half planes. This is true for diametrically opposite points by virtue of (1.4), and for other pairs of points on the two half planes by virtue of Lemma 1.1, bearing in mind that $f(c, r) > 0$ on Π_{m-2} for $r \neq 0$. There accordingly exists at least one of the half planes μ_{m-1}, say ν_{m-1}, on the closure of which $f(c, r) > 0$, $r \neq 0$. Let θ be the angle which μ_{m-1} makes with ν_{m-1} measuring θ in a definite sense, with $0 \leqq \theta < 2\pi$.

If $f(c, r) = 0$ at any point r other than $r = 0$, there will be a half plane μ_{m-1} with a maximum θ for half planes on which $f(c, r) = 0$ at some non-null point r. Let μ^*_{m-1} designate this half plane. On the half plane opposite μ^*_{m-1}, $f(c, r) \neq 0$ for $|r| = 1$. Hence a half plane for which θ is slightly larger than on μ^*_{m-1} will determine an $(m - 1)$-plane Π_{m-1} satisfying the lemma.

LEMMA 1.3. *If f is non-negative in r at $x = c$, and if (1.4) holds, then f is P.S.N. at $x = c$.*

By virtue of the preceding lemma there exists an $(n - 1)$-plane $A \cdot r = 0$ such that $f(c, r) > 0$ when $A \cdot r \geqq 0$ and $r \neq 0$. Let k be an arbitrary non-negative constant. Then

$$f(c, r) - kA \cdot r > 0 \tag{1.5}$$

when $A \cdot r < 0$, since $f \geqq 0$. But the left member of (1.5) is positive when $A \cdot r \geqq 0$, $k = 0$, and $|r| = 1$, in fact is positive and bounded from zero. Hence

the left member of (1.5) is positive when $A \cdot r \geqq 0$, $r \neq 0$, and k is positive and sufficiently small. For such a k (1.5) holds for each $r \neq 0$, and the lemma follows.

THEOREM 1.1. *The condition that f be P.S.N. at $x = c$ is equivalent to the condition* (1.4).

By virtue of the convexity of $f(c, r)$ there exists a vector a such that $f(c, r) \geqq a \cdot r$. Setting

$$\phi(x, r) \equiv f(x, r) - a \cdot r \tag{1.6}$$

we see that ϕ is convex and non-negative at $x = c$. If (1.4) holds then

$$\phi(c, r) + \phi(c, -r) > 0 \qquad (r \neq 0)$$

and we can apply the preceding lemma to ϕ to conclude that ϕ is P.S.N. at $x = c$. It follows that f is P.S.N. at $x = c$.

On the other hand suppose that f is P.S.N. at $x = c$ so that $f(c, r) > b \cdot r$ for $r \neq 0$. Then

$$f(c, r) + f(c, -r) > b \cdot r - b \cdot r = 0 \qquad (r \neq 0)$$

and the proof of the theorem is complete.

We shall say that f is *regularly convex* at c if (1.2) holds as stated, with a in (1.2) *uniquely* determined by (c, r_0) for each $r_0 \neq 0$.

LEMMA 1.4. *If f is regularly convex and non-negative at $x = c$, and if $f(c, r)$ vanishes identically on some line L through the origin in the space R, then $f(c, r)$ vanishes identically in r.*

Let r_0 be a point $r \neq 0$, not on L. Then, using (1.3),

$$f(c, r) \geqq a \cdot (r - r_0) + f(c, r_0) = a \cdot r$$

for a unique constant vector a and for every r. In particular if p is a point r on L not 0, $0 \geqq a \cdot p$. This is possible on all of L only if $a \cdot p = 0$. But $f(c, p) = 0$ by hypothesis, so that

$$f(c, r) \geqq a \cdot r = a \cdot (r - p) + f(c, p) \tag{1.7}$$

It is trivial that (1.7) holds with a replaced by 0. Since f is regularly convex at $x = c$ (1.7) can hold for but one constant a so that $a = 0$. Finally $f(c, r_0) = a \cdot r_0 = 0$, so that $f(c, r)$ vanishes identically in r.

LEMMA 1.5. *If f is regularly convex and non-negative in r at $x = c$ and $f(c, r) \not\equiv 0$ in r, then f is P.S.N. at $x = c$.*

Under the hypotheses of the lemma it follows from Lemma 1.4 that f vanishes identically on no line through the origin in the spaee R, and hence condition (1.4) holds. We conclude from Lemma 1.3 that f is P.S.N. at $x = c$.

THEOREM 1.2. *If f is regularly convex at $x = c$ the condition that f be P.S.N. at $x = c$ is equivalent to the condition that f be non-linear in r at $x = c$.*

This theorem follows upon using (1.6) much as in the proof of Theorem 1.1, Lemma 1.5 replacing Lemma 1.3.

Examples will show that Theorem 1.2 is false if the condition of regular convexity be replaced by convexity.

2. The lower semi-continuity of J

To the conditions I and II of homogeneity and convexity imposed on f in §1 we now add the following two conditions

III. *The function f shall be bounded for $|r| = 1$*

IV. *The function f shall be lower semi-continuous in x and r.*

We admit curves g of finite length and shall use representations $x(t)$, $a \leqq t \leqq b$, of g in which $x^{(i)}(t)$, $i = 1, \cdots, n$, is absolutely continuous in t. It follows from a theorem in Carathéodory [1] p. 377 that $f[x(t), \dot{x}(t)]$ is a measurable function[1] of t. Let $x = \varphi(s)$ be the representation of g in terms of arc length with $0 \leqq s \leqq s_0$. The function $f(\varphi, \dot{\varphi})$ is measurable and bounded. As a function of t, $s(t)$ is absolutely continuous. We thus have

$$\int_0^{s_0} f(\varphi, \dot{\varphi})\, ds = \int_a^b f(\varphi, \dot{\varphi})\dot{s}\, dt.$$

Upon using the homogeneity of f the latter integral takes the form

$$\int_a^b f[x(t), \dot{x}(t)]\, dt. \tag{2.1}$$

The integral (2.1) is thus independent of the particular representation $x(t)$ of g which is used. We denote the integral (2.1) by $J(g)$.[2]

We shall make use of the following lemma from McShane [2], p. 603.

LEMMA 2.1. *If f is convex in r and conditioned as in I, II, III and IV, then for each constant $k > 0$ there exists a sequence of functions $\varphi_\mu(x, r)$ which for x on A and $|r| \leqq k$ are convex in r for fixed x, of class[3] C' in (x, r), and such that for (x, r) fixed $\varphi_\mu(x, r)$ converges to $f(x, r)$ as μ becomes infinite, and $\varphi_1(x, r) < \varphi_2(x, r) < \cdots$.*

Let k be a positive constant and $x_m(t)$, $m = 0, 1, 2, \cdots, 0 \leqq t \leqq 1$ be a sequence of admissible representations of curves on A, with $|\dot{x}_m| \leqq k$ for almost all values of t and with $x_m(t)$ converging uniformly to $x_0(t)$ as m becomes infinite. Let $\varphi(x, r)$ be any one of the functions $\varphi_\mu(x, r)$. We shall show that

$$\varliminf_{m=\infty} \int_0^1 \varphi(x_m, \dot{x}_m)\, dt \geqq \int_0^1 \varphi(x_0, \dot{x}_0)\, dt. \tag{2.2}$$

---

[1] The derivative $\dot{x}(t)$ exists almost everywhere. In a Lebesgue integral the integrand can be undefined on a set E of measure 0. Or one can set such an integrand equal to an arbitrary constant on the set E.

[2] If g is a non-rectifiable curve we could define $J(g)$ as the inferior limit of $J(C)$ for rectifiable curves C as C converges in the sense of Fréchet to g. If f is P.S.N. at points on g we shall see in §3 that $J(g)$ would then be infinite. Inasmuch as f will be assumed P.S.N. in our final theorems we do not find it useful to include non-rectifiable curves in the domain of definition of $J(g)$.

[3] A function will be understood to be of class C' on a subset M of a euclidean space if it is of class C' on some region containing M.

Since $\varphi(x, r)$ is convex in r

$$\phi(x_0, \dot{x}_m) - \varphi^0 \geqq \frac{\partial \varphi^0}{\partial r} \cdot (\dot{x}_m - \dot{x}_0) \tag{2.3}$$

where the superscript indicates evaluation for $(x, r) = (x_0, \dot{x}_0)$. The t-integral of the right member of (2.3) tends to 0 with $1/m$ by virtue of a theorem in Hobson [1], §279. Accordingly

$$\varliminf_{m \to \infty} \int_0^1 \varphi(x_0, \dot{x}_m)\, dt \geqq \int_0^1 \varphi^0\, dt. \tag{2.4}$$

But the left members of (2.2) and (2.4) are equal by virtue of the uniform continuity of $\varphi(x, r)$ for x on A and $|r| \leqq k$. Relation (2.2) then follows as a consequence of this fact and of (2.4).

THEOREM 2.1. *The integral $J(g)$ is a lower semi-continuous function of g on any class of admissible curves of bounded length.*

Let g_m, $m = 1, 2, \cdots$ be a sequence of curves on A, converging in the sense of Fréchet to g_0, with lengths $L(g_m)$ at most a constant independent of m. In accordance with a lemma of McShane [1], p. 10, the curves g_m can be given representations $x_m(t)$, $0 \leqq t \leqq 1$, such that $|\dot{x}| \leqq k$ for some positive constant k, and for almost all values of t on the interval [0, 1], while $x_m(t)$ converges uniformly to $x_0(t)$. Corresponding to the constant k let $\varphi_\mu(x, r)$ be the sequence of functions affirmed to exist in Lemma 2.1. Regarded as a function of the representation $x_m(t)$ the integral

$$I_\mu(x_m) \equiv \int_0^1 \varphi_\mu(x_m, \dot{x}_m)\, dt$$

is lower semi-continuous at x_0 in accordance with (2.2). But as μ becomes infinite $I_\mu(x_m)$ tends, without decreasing, to $J(g_m)$. Hence $J(g)$ is lower semi-continuous at g_0, and the proof is complete.

3. Conditions for bounded-compactness of $J(g)$

Let c be an arbitrary constant. If for each c the set of admissible curves on which $J(g) \leqq c$ is compact, we shall say that J is *boundedly compact*. If $L(g)$ is the length of g, the set of curves for which $L(g) \leqq c$ is compact, as is well known. We accordingly seek conditions on f and J under which any class of admissible curves for which J is bounded is a class for which L is bounded.

Let $x_m(t)$, $m = 1, 2, \cdots$ be a sequence S of curve representations in which t is the arc length with $0 \leqq t \leqq a_m$. In case a_m becomes infinite with m, $x(t)$ will be called a *pseudo limit* of S if $x(t)$ is defined and absolutely continuous for $0 \leqq t < \infty$, and if there exists a subsequence $\zeta_r(t)$, $r = 1, 2, \cdots$, of the sequence $x_m(t)$ such that on each finite sub-interval for t, $\zeta_r(t)$ converges uniformly to $x(t)$. In this definition we are concerned with representations rather than curves. In the "pseudo-limit" $x(t)$, t is not necessarily the arc length of the curve $x = x(t)$.

LEMMA 3.1. *Corresponding to the sequence of representations S at least one subsequence converges to a "pseudo-limit" $x(t)$.*

In accordance with Ascoli's theorem (see Tonelli [1], p. 78) there exists a subsequence S_1 of S of the form $\bar{x}_n(t)$, $n = 1, 2, \cdots$, such that $\bar{x}_n(t)$ converges uniformly for $0 \leqq t < 1$ to an absolutely continuous function $X_1(t)$. We proceed inductively, assuming that S_{m-1} is a well defined subsequence of S_{m-2}, $m = 3, 4, \cdots$. With Ascoli we infer the existence of a subsequence of S_{m-1} of vector functions converging uniformly, for $m - 1 \leqq t < m$, to an absolutely continuous function $X_m(t)$. We define $X(t)$ for $0 \leqq t < \infty$ by setting

$$X(t) \equiv X_m(t) \qquad (m - 1 \leqq t < m, m = 1, 2, \cdots).$$

Let S^* be the subsequence $\zeta_m(t)$ of S in which $\zeta_m(t)$ is the m^{th} representation in S_m. It is clear that the "diagonal sequence" $\zeta_m(t)$ converges uniformly to $X(t)$ on each finite sub-interval for t, and that $X(t)$ is a "pseudo-limit" of S.

LEMMA 3.2. *Let C be a curve, not necessarily rectifiable, at each point of which f is P.S.N. Let λ be a positive constant. There exist positive constants δ and η so small, that when a rectifiable curve g of length λ lies on the δ-neighborhood of some point of C, then $J(g) > \eta$.*

Suppose the lemma false. There will then exist a sequence g_n of arcs of length λ converging uniformly in point-wise fashion to some point x_0 of C, while $\underline{\lim}\, J(g_n) \leqq 0$. Since f is P.S.N. at x_0 there exists a constant vector b such that

$$f(x, u) - b \cdot u > k > 0, \qquad k \text{ constant}, \tag{3.1}$$

for all vectors u of unit length, and for all points x on some neighborhood N of x_0. Let n be so large that g_n is on N. Let $x_n(s)$ be the representation of g_n in terms of arc length. It follows from (3.1) that

$$J(g_n) > \int_0^\lambda b \cdot x_n \, ds + k\lambda. \tag{3.2}$$

But the integral on the right of (3.2) converges to 0 as g_n converges point-wise to x_0. Since $k > 0$, $\underline{\lim}\, J(g_n)$ is positive. From this contradiction we infer the truth of the lemma.

THEOREM 3.1. *If a sequence of admissible curves C_m converges to a curve C, not necessarily rectifiable, and if f is P.S.N. at each point of C then $J(C_m)$ becomes infinite with $L(C_m)$.*[4]

We apply the preceding lemma taking $\lambda = 1$ and using the constants δ and η of the preceding lemma. If C be divided into successive arcs in any way, the number of these arcs with diameters $d > \delta/3$ is bounded by some integer M. Let C_m be divided into successive arcs with $L \leqq 1$ on the last arc, and $L = 1$ on the other arcs. Suppose m is so large that C and C_m admit a homeomorphism T_m in which corresponding points have a distance less than $\delta/3$.

[4] The condition that $f(x, r)$ be convex in r is not used in the proof of this theorem or of Lemma 3.2.

If the T_m-map on C of a unit arc h of C_m has a diameter $d \leqq \delta/3$, it is clear that h is within a distance δ of some point of C. At most M of the unit arcs of C_m can accordingly fail to be within a distance δ of some point of C. The remaining unit arcs of C_m each contribute at least η to $J(C_m)$. As $L(C_m)$ becomes infinite $J(C_m)$ accordingly becomes infinite and the proof of the theorem is complete.

If the admissible arcs g on which $J(g) \leqq 0$ are bounded in length by a constant λ, J will be said to satisfy the *condition of Hahn*.

In the condition of Hahn we shall always take $\lambda > 0$. If J satisfies the condition of Hahn, $J(C)$ is bounded below regardless of the length of C. In fact, if B is the absolute minimum of $f(x, r)$ for x on A and $|r| = 1$, then

$$J(C) \geqq \min (0, B\lambda).$$

An admissible curve whose length is a multiple of λ will be called a *λ-arc*. By an *initial sub-arc* of a curve h will be meant a sub-arc of h with its initial point the initial point of h. The value of J on each initial λ-arc of a λ-arc C lies between 0 and $J(C)$ inclusive. If H_m is composed of a sequence $h_1, \cdots, h_m$ of λ-arcs then

$$J(H_m) = J(h_1) + \cdots + J(h_m),$$

none of the terms in the sum being negative. If $J(H_m)$ is bounded independently of m, and m becomes infinite, then for sufficiently large values of m, H_m will possess a λ-sub-arc with an arbitrarily great length and an arbitrarily small value of J.

THEOREM 3.2. *If f is P.S.N. at each point x on A, and if J satisfies the condition of Hahn, then L is bounded on any class of curves on which J is bounded.*

For earlier theorems of this type, see Hahn [1], Graves [1], McShane [4], and Tonelli [2].

If the theorem is false there exists a sequence C_m, $m = 1, 2, \cdots$, of λ-arcs on which J is at most a positive constant M, but on which L becomes infinite with m. In the set of λ-sub-arcs of the arcs C_m, there accordingly exists a sequence g_μ, $\mu = 1, 2, \cdots$, of λ-arcs on which L becomes infinite with μ, and on which J tends to 0 with $1/\mu$. By virtue of Lemma 3.1 a suitably chosen subsequence h_p, $p = 1, 2, \cdots$, of the sequence g_μ possesses a "pseudo-limit" $x(t)$ with $0 \leqq t < \infty$. It follows from the lower semi-continuity of J that on each sub-arc of $x(t)$ on which $0 \leqq t \leqq q$ where q is a positive integer, $J \leqq 0$. By virtue of the Hahn condition the length of the curve $x(t)$ is at most λ.

As t becomes positively infinite, $x(t)$ must then tend to a limit point c, and with the addition of c as an end point define a rectifiable curve C. By virtue of the definition of a "pseudo limit," suitably chosen initial λ-arcs k_p, $p = 1, 2, \cdots$, of the respective arcs h_p converge in the sense of Fréchet to C, while $L(k_p)$ becomes infinite with p. But $J(k_p)$ is bounded by M contrary to Theorem 3.1. We infer the truth of Theorem 3.2.

Combining this theorem with the well known theorem of Hilbert on the compactness of the curve class on which $L \leqq$ const., we have the following theorem.

THEOREM 3.3. *If f is P.S.N. at each point x of A, and if J satisfies the Hahn condition, then the class of admissible curves on which J is at most a finite constant is compact.*

Theorem 1.2 leads to the following corollary of Theorem 3.3.

COROLLARY 3.1. *If f is regularly convex and non-linear in r at each point x of A, and if J satisfies the Hahn condition, then the class of admissible curves on which J is at most a finite constant is compact.*

Recall that the general conditions on f not explicitly mentioned in this corollary are those of homogeneity, convexity in r, boundedness for $|r| = 1$, and lower semi-continuity. We are not primarily concerned with the existence of a curve minimizing J and joining two given points of A. But under our general conditions on f and the special conditions of Theorems 3.2 and 3.3, or the above Corollary, such a minimizing curve exists. The compactness and lower semi-continuity necessary for the conventional proof are immediately available.

4. Convergence in length and J-length

The hypothesis that $f(x, r)$ be convex in r for each x has various consequences which we enumerate. As we have noted $f(x, r)$ is continuous in r. To continue, let u be a unit vector. For each fixed x and r, $f(x, r + hu)$ has a right derivative as to h when $h = 0$. This derivative will be denoted by $f'(x, r, u)$. It follows from Bonneson und Fenchel [1], p. 19 that

$$f(x, r) - f(x, r - u) \leqq f'(x, r, u) \leqq f(x, r + u) - f(x, r). \tag{4.1}$$

We are assuming that f is homogeneous in the sense of (1.1) and bounded for $|r| = 1$. It follows that the extreme members of (4.1) are bounded for $|r| = 1$, $|u| = 1$, and x on A. There accordingly exists a positive constant H such that

$$|f'(x, r, u)| \leqq H \tag{4.2}$$

for $|r| = 1$, $|u| = 1$ and x on A. But it follows from the homogeneity of $f(x, r)$ that

$$f'(x, kr, u) \equiv f'(x, r, u) \qquad (k \neq 0).$$

We conclude that (4.2) holds without restriction on r. An immediate consequence of (4.2) is the following.

(a) *Under our conditions on f of homogeneity in r, convexity in r, and boundedness for $|r| = 1$, the function f is continuous in r, uniformly with respect to x and r for x on A and arbitrary r.*

To obtain the theorems on convergence in length and J-length we replace the hypothesis that f be lower semi-continuous by the following

IVa. *The function f shall be continuous in x for each fixed r.*

Our general hypotheses on f are now those of

I. Homogeneity in r

II. Convexity in r

III. Boundedness for $|r| = 1$

IVa. Continuity in x.

We shall prove the following

(b) *Hypothesis III is a consequence of I, II and IVa.*

We shall show that if I, II and IVa hold, f is bounded for $|r| \leqq k$ where k is any positive constant. On account of the compactness of the set $|r| \leqq k$ it will be sufficient to show that f is bounded for x on A and r neighboring an arbitrary point p.

We first show that f is bounded above for r neighboring p. In the space R let $r_0, \cdots, r_n$ be the vertices of a simplex E containing p in its interior. Since f is continuous in x, $f(x, r_i)$ admits an upper bound M, for x on A, and $i = 0, \cdots, n$. Because of the convexity of f in r, $f \leqq M$ for x on A and r on E Hence f is bounded above for $|r| \leqq k$.

We continue by showing that f is bounded below for r neighboring p and x on A. Referring again to Bonnesen and Fenchel [1], p. 19, (2), we have the relation

$$f(x, p + hu) - f(x, p) \geqq h[f(x, p) - f(x, p - u)] \quad (0 < h < 1)$$

For us p is fixed, u is an arbitrary unit vector and x is on A. For such variables $f(x, p)$ is bounded by virtue of IVa, and $-f(x, p - u)$ is bounded below in accordance with the results of the preceding paragraph, so that $f(x, p + hu)$ is bounded below. Hence $f(x, r)$ is bounded below for x on A and r on a neighborhood of p.

Statement (b) follows.

When I, II and IVa hold the conclusion of (a) holds. This taken with IVa implies the following

(c) *Under hypotheses I, II and IVa $f(x, r)$ is continuous in (x, r) for x on A and arbitrary r.*

If a sequence of admissible curves C_m, $m = 1, 2, \cdots$, converges to an admissible curve C_0 in the sense of Fréchet, and if $J(C_m)$ converges to $J(C_0)$, we say that C_m *converges in J-length* to C_0. When $J = L$ this defines *convergence in length.*

If s is the arc length on an admissible curve C and one sets $s = tL(C)$, the parameter t is called the *reduced arc length* on C. The reduced arc length varies from 0 to 1 inclusive. A representation $x(t)$ of C in terms of reduced arc length will be called a *reduced representation* of C. The following theorem is well known: If C_m converges in length to C_0 and $x_m(t)$ and $x_0(t)$ are reduced representations of C_m and C_0 respectively then

$$\lim_{m=\infty} \int_0^1 |\dot{x}_m - \dot{x}_0| \, dt = 0 \tag{4.3}$$

and $x_m(t)$ converges uniformly to $x_0(t)$. (See McShane [3], p. 51.) Earlier results of the type of (4.3) are in Tonelli [1], p. 186, Adams and Lewy [1], p. 24 and A. Morse [1], p. 72.)

THEOREM 4.1. *Under conditions I, II and IVa on f, convergence in length to an admissible curve C_0 implies convergence in J-length to C_0.*

Let C_m, $m = 1, 2, \cdots$, be a sequence of admissible curves converging in length to C_0. Let $x_m(t)$, $m = 0, 1, \cdots$, be the reduced representation of C_m. We have $| \dot{x}_m | = L(C_m)$ for almost all values of t, $0 \leqq t \leqq 1$, so that $| \dot{x}_m |$ is bounded for almost all values of t.

To establish the theorem one notes that

$$
(4.3)\qquad \begin{aligned} J(C_m) - J(C_0) = \int_0^1 [f(x_m, \dot{x}_m) - f(x_0, \dot{x}_m)]\, dt \\ + \int_0^1 [f(x_0, \dot{x}_m) - f(x_0, \dot{x}_0)]\, dt. \end{aligned}
$$

The first integral on the right of (4.3) tends to 0 with $1/m$ by virtue of the uniform continuity of $f(x, r)$ for $| r |$ bounded. It follows from (4.2) that

$$
| f(x_0, \dot{x}_m) - f(x_0, \dot{x}_0) | \leqq H | \dot{x}_m - \dot{x}_0 |
$$

so that the second integral in (4.3) is at most

$$
(4.4)\qquad H \int_0^1 | \dot{x}_m - \dot{x}_0 |\, dt
$$

in absolute value. As we have noted before the integral (4.4) tends to 0 with $1/m$ as C_m converges in length to C_0. Thus $J(C_m) - J(C_0)$ converges to 0 with $1/m$ and the proof of the theorem is complete.

Let IIa denote *the condition on f of regular convexity*. Condition IIa implies that at (x_0, r_0) the "figurative" $z = f(x_0, r)$ has but one "supporting" n-plane

$$
(4.5)\qquad z - z_0 = a(x_0, r_0)\cdot(r - r_0) \quad r_0 \neq 0 \;\; [z_0 = f(x_0, r_0)]
$$

we shall prove the following lemma.

LEMMA 4.1. *Under the conditions I, IIa, and IVa on f, f_r exists and is a continuous function of (x, r) $r \neq 0$.*

Let $A(x_0, r_0)$ be a unit vector on the space (r, z) defining the direction of the normal to the supporting plane (4.5) taken with a positive z-component. As (x, r) tends to (x_0, r_0), $A(x, r)$ tends to $A(x_0, r_0)$. For if (x, r) tends to (x_0, r_0) through a discrete set, any cluster value A_0 of the corresponding sequence of vectors $A(x, r)$ defines a normal to an n-plane supporting the figurative at (x_0, r_0). Under condition IIa, it follows that $A_0 = A(x_0, r_0)$. Thus $A(x, r)$ is continuous in (x, r). Since each supporting n-plane is of the form (4.5) at each point (x, r) it follows that the component $A^{(n+1)}(x, r) \neq 0$. Finally

$$
a^{(i)}(x, r) = \frac{A^{(i)}(x, r)}{A^{(n+1)}(x, r)} \qquad (i = 1, \cdots, n)
$$

and we conclude that $a(x, r)$ is continuous in (x, r).

To prove the lemma it remains to establish that $f_r(x, r)$ exists and equals $a(x, r)$.

To that end let C be the convex curve in the space R cut out of the figurative $z = f(x_0, r)$ by the 2-plane Π through $r = r_0$ on which z and $r^{(i)}$ alone vary.

The curve C is supported at the point (x_0, r_0) by a unique line L. For through every line L in Π supporting C at (x_0, r_0) passes an n-plane supporting the figurative at (x_0, r_0). If u is a unit vector in the direction of the $r^{(i)}$ axis it follows that

$$f'(x_0, r_0, u) = -f'(x_0, r_0, -u)$$

so that $f_{r^{(i)}}(x_0, r_0)$ exists and equals the directional derivative $f'(x_0, r_0, u)$.

Finally this derivative equals $a^{(i)}(x_0, r_0)$. For the 2-plane Π intersects the n-plane (4.5) in a line L supporting C at (x_0, r_0) with a slope of $a^{(i)}(x_0, r_0)$ with respect to the $r^{(i)}$ axis. It follows that

$$a^{(i)}(x_0, r_0) = f_{r^{(i)}}(x_0, r_0)$$

and the proof of the lemma is complete.

We seek conditions on f under which convergence in J-length to an admissible curve C_0 implies convergence in length to C_0. We begin with an extension of a lemma of McShane, [1], p. 9, making use of the Weierstrass E-function $E(x, r, q)$.

LEMMA 4.2. *Let C_m, $m = 1, 2, \cdots$ be a sequence of curves with reduced representations $x_m(t)$ and lengths at most M, such that $x_m(t)$ converges uniformly to a representation $x_0(t)$ defining a curve C_0. Among values of t at which $\dot{x}_0$ exists let ω be the set at which $\dot{x}_0 = 0$ and let σ be the residual set. Let $b(t)$ be an arbitrary bounded measurable vector function of t.*
Then[5]

$$\begin{aligned}\underline{\lim}_{m=\infty} [J(C_m) - J(C_0)] \geqq \underline{\lim}_{m=\infty} \int_\sigma E(x_0, \dot{x}_0, \dot{x}_m)\, dt \\ + \underline{\lim}_{m=\infty} \int_\omega [f(x_0, \dot{x}_m) - b(t)\cdot\dot{x}_m]\, dt,\end{aligned} \tag{4.6}$$

the equality prevailing when $m(\omega) = 0$.

By elementary additions and subtractions of the integrands concerned on the separate sets σ and ω we find that

$$\begin{aligned}\int_\sigma & [f(x_m, \dot{x}_m) - f(x_0, \dot{x}_0)]\, dt \\ &= \int_{\sigma+\omega} [f(x_m. \dot{x}_m) - f(x_0, \dot{x}_m)]\, dt + \int_\sigma E(x_0, \dot{x}_0, \dot{x}_m)\, dt \\ &\quad + \int_\omega [f(x_0, \dot{x}_m) - b(t)\cdot\dot{x}_m]\, dt + \int_\sigma f_r(x_0, \dot{x}_0)\cdot(\dot{x}_m - \dot{x}_0)\, dt \\ &\quad + \int_\omega b(t)\cdot(\dot{x}_m - \dot{x}_0)\, dt.\end{aligned}$$

---

[5] The derivative $\dot{x}_m$ may not exist on a set E of measure 0 and on such a set the integrands in (4.5) are not defined. We understand that these integrands are set equal to an arbitrary constant on E.

The first integral on the right tends to 0 with $1/m$ by virtue of the uniform convergence of $x_m(t)$ to $x_0(t)$, the fact that $|\dot{x}_m| \leqq M$, and the uniform continuity of $f(x, r)$ over the set of variables concerned. The sum of the last two integrals tends to 0 with $1/m$ by virtue of the convergence of x_m to x_0 and a theorem in Hobson [1], §279 already cited. The lemma then follows.

LEMMA 4.3. *If the curves C_m of the preceding lemma satisfy the condition*

$$L(C_m) \geqq L(C_0) + e \qquad (m = 1, 2, \cdots) \quad (e > 0) \tag{4.7}$$

then at least one of the two following cases occur.

Case I. *The derivative $\dot{x}_0 = 0$ on a set ω of positive measure.*[6]

Case II. *There exist an integer N and positive constants η and δ such that*

$$\left| \frac{\dot{x}_m}{|\dot{x}_m|} - \frac{\dot{x}_0}{|\dot{x}_0|} \right| \geqq \eta \qquad (m > N) \tag{4.8}$$

on a set ω_m of measure exceeding δ.

We shall apply Lemma 4.2 to the function $J = L$. Assuming that Case I does not hold we shall prove that Case II must hold. When Case I does not hold, $m(\omega) = 0$. Let σ_m be the subset of σ on which $\dot{x}_m \neq 0$. According to (4.7) and Lemma 4.2 we have

$$e \leqq \underline{\lim}_{m=\infty} [L(C_m) - L(C_0)] = \underline{\lim}_{m=\infty} \int_{\sigma_m} E^*(x_0, \dot{x}_0, \dot{x}_m)\, dt$$

where E^* is the Weierstrass E-function set up for L. Developing E^* more explicitly and noting that $|\dot{x}_m| \leqq M$ almost everywhere on σ_m we have

$$e \leqq \underline{\lim}_{m=\infty} \int_{\sigma_m} \left[|\dot{x}_m| - \dot{x}_m \cdot \frac{\dot{x}_0}{|\dot{x}_0|} \right] dt \leqq M \underline{\lim}_{m=\infty} \int_{\sigma_m} \left| \frac{\dot{x}_m}{|\dot{x}_m|} - \frac{\dot{x}_0}{|\dot{x}_0|} \right| dt.$$

Hence (4.8) holds as stated and Lemma 4.3 follows.

We shall say that f is *strongly convex* at a point (x_0, r_0) at which $r_0 \neq 0$, if the convexity relation (1.2) holds with the equality excluded for each vector r differing in direction from r_0. Expressed in terms of the Weierstrass E-function this condition requires that $r_0 \neq 0$ and that $E(x_0, r_0, r) > 0$ for each vector r differing in direction from r_0.

In proving the next theorem we shall make use of the following theorem due to Lusin. If $h(t)$ is a finite measurable function defined on a bounded set g of measure exceeding a positive constant δ, there exists a subset of g of measure exceeding δ on which $h(t)$ is continuous, and hence a closed subset of g of measure exceeding δ on which $h(t)$ is continuous. See Saks [1], p. 44 for a proof.

The following theorem generalizes a theorem of Lindeberg. Cf. Tonelli [1] p. 321 and McShane [1] p. 40.

THEOREM 4.2. *Under the general conditions I, IIa, and IVa, if f is P.S.N. at each point of an admissible curve $C_0 : x = \varphi(s)$, and if f is strongly convex at*

[6] That Case I actually occurs can be shown by examples.

$[\varphi(s), \dot{\varphi}(s)]$ for almost all values of the arc length s for which $\dot{\varphi}(s) \neq 0$, then convergence in J-length to C_0 implies convergence in length to C_0.

If the theorem is false there will exist a sequence C_m, $m = 1, 2, \cdots$ of admissible curves converging in J-length to C_0 while a condition such as (4.7) holds for each m. The curves C_m are bounded in length by a positive constant M in accordance with Theorem 3.1. Hence a subsequence of these curves, which we again denote by C_m, $m = 1, 2, \cdots$, can be parameterized together with C_0, as in Lemma 4.2. See McShane [1], p. 10. Lemmas 4.2 and 4.3 then apply. We have the two cases of Lemma 4.3.

Since f is P.S.N. at each point of C_0, there exists a function $b(t)$ which is constant on each of a finite set of intervals covering [0, 1] such that for each t and $r \neq 0$

$$f[x_0(t), r] - b(t)\cdot r > k\,|\,r\,| \tag{4.9}$$

where k is a positive constant.

We make use of (4.6) taking $b(t)$ in (4.6) as the $b(t)$ of (4.9). Then

$$\varliminf_{m=\infty} [J(C_m) - J(C_0)] \geqq \varliminf_{m=\infty} \int_\sigma E(x_0, \dot{x}_0, \dot{x}_m)\,dt + m(\omega)k \varliminf_{m=\infty} L(C_m) \tag{4.10}$$

Case I. In this case $m(\omega) > 0$. Since $L(C_m)$ is bounded below by e, and $E \geqq 0$ in (4.10), we infer that $J(C_m)$ does not tend to $J(C_0)$, contrary to hypothesis. The theorem follows in Case I.

Case II. We can assume that Case I does not hold so that $m(\omega) = 0$. There exists a closed subset π_m of the set ω_m of Lemma 4.3 on which $\dot{x}_0$ is continuous and $m(\pi_m) > \delta$. The set $(x_0, \dot{x}_0, u)$ for which t is on π_m and

$$\left| u - \frac{\dot{x}_0}{|\,\dot{x}_0\,|} \right| \geqq \eta \qquad\qquad |\,u\,| = 1$$

is closed; hence on this set $E(x_0, \dot{x}_0, u) > c$ where c is a positive constant. But for almost all values of t on π_m, $|\,\dot{x}_m\,| = L(C_m)$, so that for these values of t and for $m > N$ as in (4.8),

$$E(x_0, \dot{x}_0, \dot{x}_m) > c\,|\,\dot{x}_m\,| = cL(C_m) \geqq ce \qquad (m > N).$$

It follows from (4.10), setting $m(\omega) = 0$, that $J(C_m)$ does not converge to $J(C_0)$, and from this contradiction we infer the truth of the theorem.

Institute for Advanced Study,
University of Missouri.

Bibliography

Adams, C. and Lewy, H.
1. *On convergence in length*, Duke Mathematical Journal 1 (1935), 19–26.
Bonnesen, T. und Fenchel, W.
1. *Theorie der konvexen Körper*, Springer, Berlin (1934).
Carathéodory, C.
1. *Vorlesungen über reelle Funktionen*, Teubner, Leipzig (1918).

GRAVES, L. M.

1. *On an existence theorem of the calculus of variations*, Monatshefte fur Mathematik und Physik, 39 (1932), 101–4.

   Using the hypotheses of Hahn [1], the results are extended to the n-dimensional case with the aid of two lemmas.

HAHN, H.

1. *Über ein Existenztheorem der Variations-rechnung*, Sitzungsberichte der Akademie der Wissenschaften in Wien, 134, Abt. 2A (1925), 437–47.

   The conclusion of our Theorem 3.2 is obtained (p. 439) for $n = 2$ under the hypothesis that f be of class C' and that J be positive semi-definite and positive quasi-regular, with the further conditions that $f(x, r)$ shall not vanish identically in r at any point x and that admissible arcs g on which $J(g) = 0$ are bounded in length. Our "condition of Hahn" is the generalization of this last condition as found in Tonelli [2], p. 94.

HOBSON, E. W.

1. *The theory of functions of a real variable and the theory of Fourier series*, volume II, Cambridge, (1926).

MCSHANE, E. J.

1. *Semi-continuity in the calculus of variations and absolute minima for isoperimetric problems*, Contributions to the calculus of variations, University of Chicago, (1930).
2. *Semi-continuity of integrals in the calculus of variations*, Duke Mathematical Journal 2 (1936), 597–616.
3. *Curve-space topologies associated with variational problems*, Annali della R. Scuola Normale Superiore di Pisa, Serie II, 9 (1940), 45–60.
4. *Remark concerning Mr. Graves' paper "On an existence theorem of the calculus of variations,"* Monatshefte fur Mathematik und Physik, 39 (1932), 105–6.

   It is shown that the conclusion of Lemma B of Graves [1] is a consequence of weaker hypotheses than those of Lemma A.

MENGER, K.

1. Metric methods in calculus of variations, Proceedings of the National Academy of Sciences, 23 (1937), pp. 244–250.

MORSE, A.

1. *Convergence in variation and related topics*. Transactions of the American Mathematical Society 41, (1937), 48–83.

MORSE, M.

1. *Functional topology and abstract variational theory*, Mémorial des Sciences Mathématiques, XCII, Gauthier-Villars, Paris, (1939).
2. *Functional topology*. Bulletin of the American Mathematical Society, 49 (1943), pp. 144–149.

SAKS, S.

1. *Theorie de l'integrale*, Monografje Matematyezne, Warszawa (1933).

TONELLI, L.

1. *Fondamenti di calcolo delle variazione*, Volume I, Zanichelli, Bologna, (1921).
2. *Sull 'esistenga del minimo in problemi di calcolo delle variazione*, Annali della R. Scuola Normale Superiore di Pisa, Serie II, 1 (1932), 89–90.

   On p. 94 is found a theorem similar to our Theorem 3.2 with $n = 2$. The "condition of Hahn" is assumed and J is "positive quasi-regular semi-normal." If $n = 2$ and if f have the requisite derivatives our hypotheses are equivalent to those of Tonelli.

ANNALS OF MATHEMATICS
Vol. 44, No. 3, July, 1943

THE VARIATIONAL THEORY IN THE LARGE INCLUDING THE NON-REGULAR CASE—SECOND PAPER

BY GEORGE EWING AND MARSTON MORSE

Received July 8, 1942

Introduction

In the first paper with this title the authors have reviewed and simplified the classical conditions for curve-space compactness, together with the conditions that convergence in length and convergence in J-length be equivalent. The hypothesis of convexity of the integrand $f(x, r)$ in r so effectively used by McShane [1] and others is here carried still further. Other notions such as that of a pseudo-limiting curve play an important rôle in simplifying the theorems and their proofs.

The present paper is concerned with the more novel problems of upper-reducibility and of the generalized Euler theorem or homotopy theorem. These problems belong peculiarly to the variational theory in the large. The original treatment of these problems in Morse [3] used methods which in part break down in non-regular problems. The underlying topological structure however remains essentially the same, subject to certain difficulties which have been surmounted in Morse [4] in preparation for the present paper. This paper by Morse gives the underlying definitions and the topological existence theorems for critical points.

The theorems of the present paper are first established in a Euclidean region. We turn next to a compact Riemannian manifold Σ. It would be possible to treat the problems of upper-reducibility and the homotopy theorem on Σ directly, with an appropriate use of an elaborate tensor analysis and with geodesics replacing straight lines. The necessity of doing this is avoided by the introduction of a new lemma, Lemma 7.2, on the choice of coordinate systems covering a curve. With the appropriate invariant formulation of the previously locally defined conditions of convexity, semi-normality etc., the theorems are finally stated and proved for Σ.

1. The metric spaces M, L, J

We begin with certain definitions which can be given best in terms of functions on a metric space M. Let the points of M be denoted by Greek letters α, β, γ etc., and let $\alpha\beta$ designate the distance between α and β on M. We understand that $\alpha\beta$ satisfies the usual metric axioms. We shall be concerned with two functions $J(\alpha)$ and $L(\alpha)$ of the point α of M. We suppose that $J(\alpha)$ and $L(\alpha)$ are finite, single-valued and lower semi-continuous on M. In later sections α will be identified with a curve joining two points on a Riemannian manifold, $\alpha\beta$ will be the Fréchet distance between the curves α and β while $J(\alpha)$ and $L(\alpha)$ will be integrals along α, of which $L(\alpha)$ is the length of α. We here suppose that $L \geqq 0$, and that $J(\alpha)$ is bounded below.

Beside the metric of M we shall need two other metrics, an L-metric with a distance

$$|\dot{\alpha}\beta| = \alpha\beta + |L(\alpha) - L(\beta)|$$

and a J-metric with a distance

$$(\alpha\beta) = \alpha\beta + |J(\alpha) - J(\beta)|.$$

We shall refer to the corresponding spaces as the spaces L and J. In connection with the metric spaces M, L or J, terms such as neighborhood, compact, etc., will be preceded by the letter M, L or J according to which metric is used to define these terms. If for a fixed β, $\alpha\beta$ tends to 0 and $J(\alpha)$ tends to $J(\beta)$, we say that α *converges in J-length* to β. The term *convergence in L-length* (or simply *convergence in length*) is thereby defined, taking $J = L$.

Let E be a subset of L. Let I_τ be the interval $0 \leqq \tau \leqq a$, $a > 0$. By $E \times I_\tau$ we shall mean the product of E and I_τ and shall assign the usual metric to this product space. We shall admit deformations D of E which replace a point α found on E at the time $\tau = 0$, by a point $\varphi(\alpha, \tau)$ on L at the time τ, $0 \leqq \tau \leqq a$. We also require the following

(a). *The point function $\varphi(\alpha, \tau)$ shall map $E \times I_\tau$ continuously into L.*

(b). *For each fixed α on E, $\varphi(\alpha, \tau)$ shall map I_τ continuously into M, uniformly with respect to α and τ.*

We term D a *weak J-deformation* if for each fixed α on E, and for $\varphi = \varphi(\alpha, \tau)$, $J(\alpha) - J(\varphi)$ is negative for no value of τ on I_τ. A weak J-deformation is termed *proper* if the above difference $J(\alpha) - J(\varphi)$ is positive and bounded from 0 uniformly for α on E, if τ on I_τ is bounded from 0.

Let J^b denote the subset of points of M for which $J \leqq b$. We say that J is *upper-reducible at α and on J^b* if $b > J(\alpha)$ and if corresponding to an arbitrary constant a with $b > a > J(\alpha)$ there exists a weak J-deformation of some M-neighborhood N of α relative to J^b which is a proper deformation of $N \cdot (J^b - J^a)$. We term J *upper-reducible at α* if J is upper-reducible at α on each set J^b. In interpreting these definitions we understand that the null deformation is a weak J-deformation, and that any weak J-deformation is a proper deformation of an empty set.

As is well known the integrals of the calculus of variations are not in general upper semi-continuous. They are however upper-reducible under very general conditions as we shall see, and this upper-reducibility fills the gap caused by the lack of upper semi-continuity. The earlier proofs of upper-reducibility in the regular case made use of elementary extremals. Since these elementary extremals are not in general available in the non-regular case, a different type of proof is necessary. The proof given in the next section is more closely related to the proof of the upper-reducibility of the Douglas-Dirichlet integral than to the earlier proof for regular simple integrals. See Morse and Tompkins [1].

2. A deformation problem

The parameters in each local representation of our Riemannian manifold Σ will be coordinates $(x^{(1)}, \cdots, x^{(n)})$ on a bounded region S of Cartesian n-space. In this local system we shall use the notation of vector analysis[1] letting x represent the vector whose coordinates are $(x^{(1)}, \cdots, x^{(n)})$. It will simplify matters and cause no loss of generality if S is assumed convex.

We shall admit curves on S with vector representations $x(t)$, $0 \leqq t \leqq t_0$, in which each $x^{(j)}(t)$ is absolutely continuous. The t-derivative of $x(t)$ will be denoted by $\dot{x}(t)$. A particularly useful representation is that in terms of reduced length. Let α be a curve with length $L(\alpha)$. To obtain a representation of α in terms of reduced length t one sets $s = tL(\alpha)$, where s is the arc length on α measured from the initial point of α. The parameter t then ranges from 0 to 1 inclusive. If $x(t)$ is the reduced representation of α, $|\dot{x}| = L(\alpha)$ for almost all values of t.

We have introduced various distances between curves. To state our results in the briefest fashion it will be convenient to introduce a *distance between two representations* $x(t)$ and $z(t)$ of two curves given with the same interval $[0, h]$ for t. This distance will be defined by the relation

$$d[x(t), z(t)] = \max |x(t) - z(t)| \qquad (0 \leqq t \leqq h)$$

If for a fixed $z(t)$, $d(x, z)$ tends to 0, $x(t)$ converges uniformly to $z(t)$.

As a matter of permanent notation let β be a fixed rectifiable curve and α a variable curve, both of positive length. Let $x(t)$ and $y(t)$ respectively be admissible representations of α and β, with $0 \leqq t \leqq 1$. We shall have occasion to consider a deformation in which β is fixed and α is replaced at the time τ by a curve with the representation

$$\eta_\alpha^\tau(t) = x(t) + \tau[y(t) - x(t)] \qquad (0 \leqq \tau \leqq 1) \tag{2.1}$$

It is clear that such a deformation of α depends upon the representation $x(t)$ and $y(t)$, and not merely upon α and β as points of M or L. We would like to choose representations of α and β such that the following three conditions are fulfilled.

(A) *The deformation* (2.1) *is admissible in the sense of* (a) *and* (b) *of* §1.

(B) *For any class of curves α for which $L(\alpha)$ is bounded and $\alpha\beta$ sufficiently small, $|\dot{\eta}_\alpha^\tau|$ is bounded independently of t, τ, and α.*

(C) *As $\alpha\beta$ tends to* 0, *$d[y(t), \eta_\alpha^\tau(t)]$ tends to* 0 *uniformly with respect* to τ.

We have been unable to find representations of α and β such that (A), (B) and (C) are satisfied simultaneously. If α and β are represented in terms of reduced length one could show that (A) and (B) are satisfied but that (C) fails in general. To meet this difficulty we modify our approach as follows. We seek not one representation $y(t)$ of β but many, in fact a representation $y_\alpha(t)$ of β which is

---

[1] When dealing with Σ in §7 and thereafter we shall use the notation of tensor analysis.

determined by α and β together. We are thus concerned with a deformation of the form

$$x_\alpha^\tau(t) = x(t) + \tau[y_\alpha(t) - x(t)]. \tag{2.2}$$

The representation $x(t)$ of α shall be in terms of reduced length. Because we have not been able to satisfy (C) as well as (A) and (B) we shall abandon the attempt to find a single deformation satisfying (A), (B) and (C) and seek a sequence of deformations $D(m, \beta)$, $m = 1, 2, \cdots$ each of which satisfy (A) and (B) and which are such that the following is true.

(D) *As* $\alpha\beta$ *and* $1/m$ *tend to* 0, $d(y_\alpha, x_\alpha^\tau)$ *tends to* 0 *uniformly with respect* to τ.

Our representation $y_\alpha(t)$ of β will depend both on α and m, so that in $D(m, \beta)$, $x_\alpha^\tau(t)$ will depend on the parameter m. The parameter m will ordinarily be held fast and so will not be explicitly indicated.

3. The (α, m)-representation $y_\alpha(t)$ of β

In accordance with the preceding section α is a variable curve with a reduced representation $x(t)$, and β is a fixed curve whose representation $y_\alpha(t)$ will depend on α and on an integer m. As an aid in defining $y_\alpha(t)$ we make use of earlier results of Morse [2] whereby the point x of S on α can be represented in terms of a μ-parameter[2] as follows. Let I_μ represent the interval $0 \leqq \mu \leqq 1$. In the μ-representation of α, x is given by a function

$$x = X(\mu, \alpha) \tag{3.1}$$

which maps $I_\mu \times M$ continuously into S. For a fixed α, $X(\mu, \alpha)$ is a representation of α, and as stated $X(\mu, \alpha)$ varies continuously on S with μ on I_μ and α on M.

We make use of the representations $X(\mu, \alpha)$ and $X(\mu, \beta)$ of α and β respectively to divide α and β into m successive arcs α_i and β_i respectively $(i = 1, \cdots m)$ on each of which $\Delta\mu = 1/m$. Recalling that t represents the reduced length on α, let t_{i-1}^α and t_i^α be the values of t on α at the end points of α_i. We parameterize β_i as follows. The parameter on β_i will be denoted by t and have the values t_{i-1}^α and t_i^α at the end points of β_i. At an inner point p of β_i the parameter t shall be such that $t - t_{i-1}^\alpha$ is proportional to the arc length on β_i measured from its initial point to p. The resulting representation of β is our (α, m)-representation $y_\alpha(t)$ of β. Let L^+ denote the subset of curves of L of positive length. We are supposing that α lies on L^+.

We shall recall certain properties of convergence in length. See McShane [2], p. 51. If an arc with a reduced representation $\eta(t)$ converges in length to an arc with a reduced representation $\zeta(t)$, then $\eta(t)$ converges uniformly to $\zeta(t)$ and $\dot\eta(t)$ converges to $\dot\zeta(t)$ in the mean (understood of the first order). If a curve α converges in length to a curve γ, the length of any subarc tends to the length of any subarc of γ to which it converges in the sense of Fréchet as $\alpha\gamma$

---

[2] Actually the present parameter is *reduced μ-length,* bearing the same relation to the parameter μ of Morse [3] that reduced length bears to length.

tends to 0. With this understood we see that, for m and β fixed, t_i^α and $L(\alpha_i)$ are continuous functions of α on L. Two first properties of the (α, m)-representation $y_\alpha(t)$ of the fixed curve β result as follows.

(i) *For a fixed m, $y_\alpha(t)$ is continuous in t and α, for t on $[0, 1]$ and α on L.*

(ii) *For a fixed m and curves α and γ on L^+, $y_\alpha(t)$ converges uniformly to $y_\gamma(t)$ and $\dot{y}_\alpha(t)$ converges in the mean to $\dot{y}_\gamma(t)$ as α converges in length to γ.*

For almost all t on the interval $[t_{i-1}^\alpha, t_i^\alpha]$

$$| \dot{y}_\alpha(t) | = \frac{L(\beta_i)}{L(\alpha_i)} L(\alpha). \tag{3.2}$$

If then α is restricted to a class L^κ of curves of L^+ whose lengths are at most κ then for almost all t

$$| \dot{y}_\alpha(t) | \leqq \kappa \frac{L(\beta_i)}{L(\alpha_i)} \qquad [L(\alpha) \neq 0]. \tag{3.3}$$

For each m there exists an M-neighborhood N_m of β so small that when α is on N_m

$$L(\alpha_i) > \frac{L(\beta_i)}{2} \tag{3.4}$$

by virtue of the lower semi-continuity of $L(\alpha_i)$. With the aid of (3.3) and (3.4) we then have the result.

(iii) *Corresponding to the curves α on L^κ and the integer m there exists an M-neighborhood N_m of β such that for almost all t*

$$| \dot{y}_\alpha(t)| < 2\kappa \qquad [\alpha \text{ on } (L^\kappa \cdot N_m)]. \tag{3.5}$$

As $\alpha\beta$ tends to 0 the euclidean distance between the points of α and β respectively bearing the parameter t_i^α tends to 0 uniformly for all $m > 0$ and $i = 1, \cdots, m$. Moreover the euclidean diameters of the arcs α_i and β_i used in the definition of $y_\alpha(t)$ tend to 0 with $\alpha\beta$ and $1/m$. We thus obtain a final property of $y_\alpha(t)$.

(iv) *The distance $d[x(t), y_\alpha(t)]$ tends to 0 as $\alpha\beta$ and $1/m$ both tend to 0.*

4. The deformation $D(m, \beta)$

We can now define a deformation $D(m, \beta)$ satisfying conditions (A), (B) and (D) of §2. In this deformation β is a fixed curve with a variable representation. The time τ varies on the interval $[0, 1]$. The curve α to be deformed has a positive length and a reduced representation $x(t)$. For each fixed positive integer m we use the (α, m)-representation $y_\alpha(t)$ of β. At the time τ the image $\varphi(\alpha, \tau)$ of α under $D(m, \beta)$ shall have the representation

$$x_\alpha^\tau(t) = x(t) + \tau[y_\alpha(t) - x(t)] \qquad (0 \leqq \tau \leqq 1) \tag{4.0}$$

The condition (a). We begin by showing that $\varphi(\alpha, \tau)$ satisfies (a) of §1.

To that end let α and α_0 be curves of L^+, and let τ and τ_0 be values of τ on

[0, 1]. Let $x(t)$ and $x_0(t)$ be reduced representation of α and α_0 respectively. Set

$$\tau - \tau_0 = \Delta\tau, \qquad x(t) - x_0(t) = \Delta x$$
$$x^\tau_\alpha(t) - x^{\tau_0}_{\alpha_0}(t) = \Delta x^\tau_\alpha, \qquad y_\alpha(t) - y_{\alpha_0}(t) = \Delta y_\alpha$$

Referring to (4.0) we obtain

$$\Delta x^\tau_\alpha = (1 - \tau_0)\Delta x + (y_\alpha - x)\Delta\tau + \tau_0\Delta y_\alpha. \tag{4.1}$$

Taking maxima as t ranges over [0, 1] we have

$$\max |\Delta x^\tau_\alpha| \leqq (1 - \tau_0) \max |\Delta x| + |\Delta\tau| \max |x - y_\alpha| + \tau_0 \max |\Delta y_\alpha|. \tag{4.2}$$

If α converges in length to α_0, $\max |\Delta x|$ and $\max |\Delta y_\alpha|$ will tend to 0, (see (i) §3), and if $|\Delta\tau|$ also tends to 0, $\max |\Delta x^\tau_\alpha|$ will tend to 0 in accordance with (4.2).

To show that $\varphi(\alpha, \tau)$ satisfies (a) it remains to show that $\varphi(\alpha, \tau)$ converges in length to $\varphi(\alpha_0, \tau_0)$ as α converges in length to α_0 and τ converges to τ_0.

The lengths of $\varphi(\alpha, \tau)$ and $\varphi(\alpha_0, \tau_0)$ have an absolute difference

$$\left| \int_0^1 (|\dot x^\tau_\alpha| - |\dot x^{\tau_0}_{\alpha_0}|)\, dt \right| \leqq \int_0^1 |\dot x^\tau_\alpha - \dot x^{\tau_0}_{\alpha_0}|\, dt = \int_0^1 |\Delta \dot x^\tau_\alpha|\, dt. \tag{4.3}$$

From (4.1) we find that

$$\Delta \dot x^\tau_\alpha = (1 - \tau_0)\Delta\dot x + (\dot y_\alpha - \dot x)\Delta\tau + \tau_0\Delta\dot y_\alpha \tag{4.4}$$

for almost all t. But in accordance with (ii) of §3, $\dot x$ and $\dot y_\alpha$ converge in the mean to $\dot x_0$ and $\dot y_{\alpha_0}$ respectively as α converges in length to α_0. If in addition $|\Delta\tau|$ converges to 0 it follows from (4.4) that the integral

$$\int_0^1 |\Delta \dot x^\tau_\alpha|\, dt$$

converges to 0. Hence $\varphi(\alpha, \tau)$ converges in length to $\varphi(\alpha_0, \tau_0)$ as α converges in length to α_0 and τ converges to τ_0. Cf. McShane [2], p. 51.

Thus condition (a) is satisfied.

The condition (b). That condition (b) of §1 is satisfied by $\varphi(\alpha, \tau)$ follows at once from (4.2) upon setting $|\Delta x|$ and $|\Delta y_\alpha|$ equal to 0 in (4.2) and observing that $|x - y_\alpha|$ is bounded since S is bounded.

Of the conditions (A), (B) and (D) of §2, (A) is thus established for $\varphi(\alpha, \tau)$. Condition (B) is satisfied as a consequence of property (iii) of $y_\alpha(t)$ in §3. To show that (D) is satisfied note that

$$d(y_\alpha, x^\tau_\alpha) = (1 - \tau) \max |x(t) - y_\alpha(t)| = (1 - \tau) d(x, y_\alpha) \tag{4.5}$$

where the maximum is for $0 \leqq t \leqq 1$. Property (iv) of §3 of y_α shows that this distance tends to 0 with $\alpha\beta$ and $1/m$. Thus (A), (B) and (D) are satisfied by the deformation $\varphi(\alpha, \tau)$.

We add the lemma.

LEMMA 4.1. *The image $\varphi(\alpha, \tau)$ of α under $D(m, \beta)$ satisfies the relation*

$$L(\varphi) \leqq L(\alpha) + \tau[L(\beta) - L(\alpha)] \leqq \max [L(\alpha), L(\beta)] \qquad (0 \leqq \tau \leqq 1). \tag{4.6}$$

We see that

$$\begin{aligned} L(\varphi) &= \int_0^1 | \dot{x}_\alpha^\tau | \, dt = \int_0^1 | (1 - \tau)\dot{x} + \tau \dot{y}_\alpha | \, dt \\ &\leqq (1 - \tau) \int_0^1 | \dot{x} | \, dt + \tau \int_0^1 | \dot{y}_\alpha | \, dt. \end{aligned} \tag{4.7}$$

Relation (4.6) follows.

Relation (4.7) shows that $L(\varphi)$ is convex in τ.

5. The upper-reducibility of J

To define our integral J we require a function $f(x, r)$ of the point x on S and of a vextor r. Let S' be a region containing the closure $\bar{S}$ of S. We suppose that f is of class[3] C' for x on S' and any $r \neq 0$. The function f shall be homogeneous in the sense that

$$f(x, kr) \equiv kf(x, r) \qquad (k \geqq 0).$$

We also suppose that f is a convex function of r. This implies that the Weierstrass E-function is never negative. Under these hypotheses $J(\alpha)$ is a continuous function of α on L and a lower semi-continuous function of α on M on any subset L^κ of M. See Morse and Ewing [1]. We shall prove that $J(\alpha)$ is upper-reducible at each curve β of L^+.

We make use of the deformation $D(m, \beta)$ for a fixed m and β. Recall that the subset of curves of positive lengths at most κ is denoted by L^κ. As a function of τ, $J(\varphi)$ is almost convex in the sense of the following lemma.

LEMMA 5.1. *If α and β belong to L^κ and $\varphi(\alpha, \tau)$ is the image of α under the deformation $D(m, \beta)$, then*

$$J(\varphi) \leqq J(\alpha) + \tau[J(\beta) - J(\alpha)] + \tau h(\alpha, \beta, m) \qquad (0 \leqq \tau \leqq 1) \tag{5.1}$$

where $h \geqq 0$ and is less than a preassigned positive constant e if m is sufficiently large and α is on a sufficiently small M-neighborhood Ω_m of β.

By virtue of the convexity of $f(x, r)$ in r,

$$J(\varphi) \leqq \int_0^1 f(x_\alpha^\tau, \dot{x}) \, dt + \tau \int_0^1 [f(x_\alpha^\tau, \dot{y}_\alpha) - f(x_\alpha^\tau, \dot{x})] \, dt. \tag{5.2}$$

The differences

$$x_\alpha^\tau(t) - x(t) = \tau(y_\alpha - x), \qquad x_\alpha^\tau(t) - y_\alpha(t) = (1 - \tau)(x - y_\alpha) \tag{5.3}$$

[3] By virtue of the homogeneity of f in r, f_x exists and is continuous even when $r = 0$.

converge to 0 uniformly with respect to t as $\alpha\beta$ and $1/m$ tend to 0 in accordance with property (D) of the deformation $D(m, \beta)$. Hence for $x = x(t)$

$$\int_0^1 f(x_\alpha^\tau, \dot{x})\, dt - \int_0^1 f(x, \dot{x})\, dt \leqq R_1 \tag{5.4}$$

where $R_1 \geqq 0$ and is of the character of τh in (5.1). Use is thereby made of the fact that $|\dot{x}| \leqq \kappa$ for almost all t, and that for $|r| \leqq \kappa$ and x on S, f_x is bounded. In the second term on the right of (5.2), $|\dot{y}_\alpha| \leqq 2\kappa$ in accordance with (iii) of §3, provided m is sufficiently large and α is on $N_m \cdot L^\kappa$. The second term is thus of the form

$$\tau[J(\beta) - J(\alpha)] + \tau h_1(\alpha, \beta, m) \tag{5.5}$$

where $|h_1|$ has the character of h in the lemma. The lemma follows.

It follows from (5.1) that $D(m, \beta)$ is a proper J-deformation of any point α of L^κ on a sufficiently small M-neighborhood of β for which

$$J(\alpha) - J(\beta) > h(\alpha, \beta, m) \tag{5.6}$$

initially. But there may be points α arbitrarily near β at which $J(\alpha) \leqq J(\beta)$, and for such points we could not affirm that $D(m, \beta)$ was even a weak J-deformation of α. We are able however to modify $D(m, \beta)$, depending upon the initial value of $J(\alpha)$, obtaining thereby a deformation $D(m, \beta, a)$ which more nearly meets our needs.

The deformation $D(m, \beta, a)$. Let a and c be constants such that

$$a > J(\beta), \qquad c = \frac{a + J(\beta)}{2}.$$

For each α on L^+ we shall now define a value $\tau(\alpha)$ of the time τ. For points α of L^+ at which $c \leqq J(\alpha) \leqq a$ let $\tau(\alpha)$ be a value of τ which divides the interval $[0, 1]$ in the same ratio as that in which $J(\alpha)$ divides the interval $[c, a]$. For $J(\alpha) > a$ we take $\tau(\alpha)$ as 1, and for $J(\alpha) < c$ we take $\tau(\alpha)$ as 0. Recalling that $J(\alpha)$ is continuous in α on L, by virtue of hypotheses at the beginning of §5 together with Theorem 4.1 of Morse and Ewing [1], we see that $\tau(\alpha)$ is also continuous on L^+. The deformation $D(m, \beta, a)$ is now obtained from $D(m, \beta)$ by deforming α as under $D(m, \beta)$ until the time $\tau(\alpha)$ is reached, and holding the image of α fast thereafter. This deformation has the following properties.

Like $D(m, \beta)$ it satisfies conditions (a) and (b) of §1 and so is admissible.

In terms of the preceding constant c we write (5.1) in the form

$$J(\varphi) - J(\alpha) \leqq \tau[J(\beta) - c + h] + \tau[c - J(\alpha)]. \tag{5.7}$$

By virtue of Lemma 5.1, if m is sufficiently large, say $m > m_1$ and α is on $\Omega_m^c \cdot L^\kappa$, where Ω_m^c is a sufficiently small M-neighborhood of β, then $h(\alpha, \beta, m)$ will be less than the positive constant $c - J(\beta)$ and (5.7) will take the form

$$J(\varphi) - J(\alpha) \leqq \tau[c - J(\alpha)] \leqq 0 \qquad (m > m_1). \tag{5.8}$$

Hence for $m > m_1$, $D(m, \beta, a)$ is a weak J-deformation of $\Omega_m^c \cdot L^\kappa$. This follows for $J(\alpha) > c$ from (5.8), and for $J(\alpha) \leqq c$ from the fact that $D(m, \beta, a)$ is the null deformation. From (5.8) we see that for $m > m_1$, $D(m, \beta, a)$, is a proper deformation of the subset of $\Omega_m^c \cdot L^\kappa$ for which $J(\alpha) \geqq a$.

To establish the upper-reducibility of J on L at a curve β we assume that L is bounded with J. Conditions on f that L be bounded with J are given in Morse and Ewing [1], Theorem 3.2. When L is bounded with J, a set J^b lies on some set $L \leqq \kappa$ for κ sufficiently large. To establish the upper-reducibility of J at a curve β of positive length we refer to the definition of upper-reducibility and choose an arbitrary constant $a > J(\beta)$ and a second arbitrary constant $b > a$. To apply Lemma 5.1 we choose κ so large that J^b is on the set $L \leqq \kappa$. We then apply the deformation $D(m, \beta, a)$ with $m > m_1$. Under this deformation $\Omega_m^c \cdot L^\kappa$ is weakly J-deformed, while $\Omega_m^c \cdot (J^b - J^a)$ suffers a proper J-deformation. Hence J is upper-reducible on J^b at β. But a and b are arbitrary constants subject to the condition $b > a > J(\beta)$. Hence J is upper-reducible on L at β, in accordance with the definition of upper-reducibility.

We thus have the following theorem.

THEOREM 5.1. *If L is bounded with J, J is upper-reducible on L at each rectifiable curve β of positive length.*[4]

6. The homotopy theorem

The function J is said to be *homotopically ordinary* at a point β of L if there exists a proper J-deformation of some J-neighborhood of β. If not homotopically ordinary at β, J will be termed *homotopically critical*. The fundamental theorem of this section is that under suitable conditions on f a point β which is homotopically critical is an "extremal". This is a semi-topological generalization of the theorem of Euler that, under suitable conditions on f, a minimizing curve is an extremal. For a minimizing curve clearly defines a homotopic critical point.

We shall need the condition that $f(x, r)$ be *strongly convex* at (x, ρ), $\rho \neq 0$. Since we are assuming that f is of class C', the required condition is that $E(x, \rho, r) \geqq 0$ for the given (x, ρ) and arbitrary r, and that it vanish only when $r = k\rho$ where k is a non-negative constant. With this understood let $x = x(s)$, $0 \leqq s \leqq s_0$, be a representation of a curve β in terms of length. A pair $[x(s), \dot{x}(s)]$ at which $\dot{x}(s) \neq 0$ will be termed an *element tangent* to β. A set of elements tangent to β will be said to include *almost all* such elements if the corresponding values of s include almost all values of s on the interval $[0, s_0]$. With this understood Theorem 4.2 of Morse and Ewing [1] takes the following form.

THEOREM 6.0. *If f is strongly convex at almost all elements tangent to β and is positive semi-normal at all points of β, then α converges in L-length to β if α converges in J-length to β.*

---

[4] The theorem holds even when $L(\beta) = 0$, as one shows by a trivial modification of the proof.

We shall make use of "variations" $\eta(t)$ of class C' for $0 \leqq t \leqq 1$, vanishing for $t = 0$ and $t = 1$.

Let β be a fixed curve of positive length and let α be a curve on an L-neighborhood of β. Let $x(t)$ be the reduced representation of α. We shall consider a deformation $\Delta(\eta, \beta)$ of an L-neighborhood of β in which α is replaced at the time τ by a curve $\psi(\alpha, \tau)$ on L with the representation

$$z_\alpha^\tau(t) = x(t) + \tau\eta(t) \qquad (0 \leqq t \leqq 1) \qquad (0 \leqq \tau \leqq e). \tag{6.1}$$

Recall that $|\dot{x}(t)| = L(\alpha)$ for almost all t.

The neighborhood N and the value of e. We shall restrict α to so small an L-neighborhood N of β that $L(\alpha)$ is bounded and bounded from 0. We take e and N so small that for α on $\bar{N}$ and τ on $[0, e]$, the image curves z_α^τ are on S, and for almost all t

$$\rho > |\dot{z}_\alpha^\tau| = |\dot{x} + \tau\dot{\eta}| > \sigma > 0 \tag{6.2}$$

where ρ and σ are positive constants.

LEMMA 6.1. *The deformation $\psi(\alpha, \tau)$ satisfies the conditions (a) and (b) of §1 and is accordingly admissible.*

To establish (a) suppose that α converges in length to α_1 and τ converges to τ_1. We have

$$|z_\alpha^\tau(t) - z_{\alpha_1}^{\tau_1}(t)| \leqq |x(t) - x_1(t)| + |\tau - \tau_1|\,|\eta(t)|. \tag{6.3}$$

That $x(t)$ converges uniformly to $x_1(t)$ as α converges in length to α_1 follows from the fact that t is the reduced length on α and α_1 respectively. From (6.3) we see then that $\psi(\alpha, \tau)$ converges in the sense of Fréchet to $\psi_1 = \psi(\alpha_1, \tau_1)$, as α converges in length to α_1 and τ converges to τ_1.

It remains to show that ψ converges in length to ψ_1. Observe that

$$\begin{aligned} |L(\psi) - L(\psi_1)| &= \left|\int_0^1 \{|\dot{x} + \tau\dot{\eta}| - |\dot{x}_1 + \tau_1\dot{\eta}|\}\,dt\right| \\ &\leqq \int_0^1 |\dot{x} - \dot{x}_1|\,dt + \int_0^1 |\tau - \tau_1|\,|\dot{\eta}|\,dt. \end{aligned} \tag{6.4}$$

As noted in §3 the integral of $|\dot{x} - \dot{x}_1|$ converges to 0 as α converges in length to α_1. Relation (6.4) then shows that ψ converges in length to ψ_1 as stated. The deformation $\psi(\alpha, \tau)$ thus satisfies (a).

Condition (b) of §1 is satisfied if $\psi(\alpha, \tau)$ maps the interval for τ continuously into M, uniformly with respect to (α, τ). That condition (b) is satisfied follows from the form of (6.1).

The first variation of J. Relation (6.2) holds for α on N and τ on [0, 3]. For such an α and τ set

$$\int_0^1 f(z_\alpha^\tau, \dot{z}_\alpha^\tau)\,dt \equiv w(\alpha, \tau). \tag{6.5}$$

Granting the possibility of differentiating under the integral sign we have

$$w_\tau(\alpha, \tau) = \int_0^1 \{f_x(z^\tau_\alpha, \dot z^\tau_\alpha)\cdot\eta + f_r(z^\tau_\alpha, \dot z^\tau_\alpha)\cdot\dot\eta\}\, dt. \tag{6.6}$$

To justify this differentiation one forms the difference quotient Q from the integrand in (6.5), assuming values τ and $\tau + \Delta\tau$. Since f is of class C' and (6.2) holds $|Q|$ is bounded for almost all t and for a sequence of values of $\Delta\tau$ converging to 0. As $\Delta\tau$ tends to 0, Q converges to the integrand of (6.6) for almost all t. It follows from the Lebesgue integration theorem that (6.6) holds as stated.

For the remainder of this section we add the hypothesis that f be of class C'' for $r \neq 0$.[5]

We refer to the constants ρ and σ of (6.2) and state the following lemma.

LEMMA 6.2. *For x on S and $|p|$ and $|q|$ on the closed interval $[\sigma, \rho]$*

$$|f_x(x, p) - f_x(x, q)| \leqq H\,|p - q| \tag{6.7}$$

$$|f_r(x, p) - f_r(x, q)| \leqq K\,|p - q| \tag{6.8}$$

where H and K are positive constants.

Relations (6.7) and (6.8) hold when $p = q$ regardless of the choice of H and K. To establish (6.7) note that

$$\frac{|f_x(x, p) - f_x(x, q)|}{|p - q|} \qquad (p \neq q) \tag{6.9}$$

is bounded for the variables admitted in the lemma if one excludes a neighborhood of the set of pairs (p, q) in which $p = q$. The set of pairs (p, p) for which $|p|$ is on the closed interval $[\sigma, \rho]$ form a closed set T. Neighboring each pair of T but excluding pairs on T the quotient (6.9) is bounded, since f is of class C''. It follows that the quotient (6.9) has a bound H for the variables admitted. Hence (6.7) holds. The proof of (6.8) is similar.

The following lemma is essential.

LEMMA 6.3. *The function $w_\tau(\alpha, \tau)$ is continuous in (α, τ) on the domain for which* (6.2) *holds, that is for α on N and τ on* $[0, e]$.

We begin by proving (i) and (ii).

(i) *The function $w_\tau(\alpha, \tau)$ is continuous in τ uniformly for α on N.*

The partial derivatives $f_x(x, r)$ and $f_r(x, r)$ appearing in (6.6) are uniformly continuous in x and r, for x on S and $|r|$ on the interval $[\sigma, \rho]$. It follows that the integrand of (6.6) is a continuous function of τ, uniformly for α on N and for almost all t on $[0, 1]$. Statement (i) follows.

(ii) *For each τ on $[0, e]$, $w_\tau(\alpha, \tau)$ is a continuous function of α on N.*

To establish (ii) let α and α_1 be points on N and set

$$u(t) \equiv x^\tau_\alpha(t), \qquad v(t) \equiv x^\tau_{\alpha_1}(t).$$

[5] Our results could be obtained with suitable changes in proof if f_x and f_r were merely subject to appropriate Lipschitz conditions.

From the difference

$$
\begin{aligned}
w_\tau(\alpha, \tau) - w_\tau(\alpha_1, \tau) = & \int_0^1 [f_x(u, \dot u) - f_x(v, \dot u)]\cdot \eta \, dt \\
& + \int_0^1 [f_x(v, \dot u) - f_x(v, \dot v)]\cdot \eta \, dt \\
& + \int_0^1 [f_r(u, \dot u) - f_r(v, \dot u)]\cdot \dot\eta \, dt \\
& + \int_0^1 [f_r(v, \dot u) - f_r(v, \dot v)]\cdot \dot\eta \, dt.
\end{aligned} \tag{6.10}
$$

Note that

$$u - v = x(t) - x_1(t). \tag{6.11}$$

This difference tends uniformly to 0 as α converges in length to α_1, since t is the reduced length on α and α_1. Moreover $| \dot v |$ and $| \dot u |$ are on the interval $[\sigma, \rho]$ of (6.2) for almost all t. It follows that the first and third integrals in (6.10) tend to 0 as α converges in length to α_1. Upon using (6.7) we see that the second integral in (6.10) is at most

$$\int_0^1 H \, | \dot u - \dot v | \, | \eta | \, dt = \int_0^1 H \, | \dot x - \dot x_1 | \, | \eta | \, dt$$

and this again tends to 0 as α converges in length to α_1. The last integral in (6.10) similarly tends to 0. Statement (ii) follows.

The lemma is an immediate consequence of (i) and (ii).

Let $y(t)$ be a reduced representation of β. Then

$$w_\tau(\beta, 0) = \int_0^1 [f_x(y, \dot y)\cdot \eta + f_r(y, \dot y)\cdot \dot\eta] \, dt \equiv I(\eta),$$

introducing $I(\eta)$. Regarded as a function of η, $I(\eta)$ is the "first variation" of J at β. We say that J has a "null first variation at β" if $I(\eta) = 0$ for all admissible variations η. We have required that η be of class C'. For our purposes it would be immaterial if we had taken $\eta(t)$ absolutely continuous. For upon using the approximation theorems of the Lebesgue theory it is not difficult to prove that under the conditions $\eta(0) = \eta(1) = 0$, a necessary and sufficient condition that $I(\eta) = 0$ for all η of class C' is that $I(\eta) = 0$ for all η which are absolutely continuous.

With this understood we term a rectifiable curve β an *extremal* if J has a null first variation at β. The fundamental theorem of this section is then the following.

THEOREM 6.1. *If β is homotopically critical and if convergence in J-length to β implies convergence in L-length to β, then β is an extremal.*

The theorem will be established in the following equivalent form. If the first

variation $I(\eta)$ of J at β is negative for some admissible variation η, and I convergence in J-length to β implies convergence in L-length to β then β is a homotopically ordinary point of J. We shall show that J is homotopically ordinary at β by exhibiting a proper J-deformation of some L-neighborhood of β. In this connection the definition of homotopically ordinary requires a J-neighborhood but under the hypotheses of the theorem every L-neighborhood of β contains a J-neighborhood of β so that an L-neighborhood may be used instead. See Morse and Ewing [1], Theorem 4.2 for conditions under which convergence in J-length implies convergence in length.

Corresponding to the given η we set up the deformation $\Delta(\eta, \beta)$. Then $w_\tau(\beta, 0) < 0$. By virtue of its continuity, $w_\tau(\alpha, \tau) <$ const. < 0, for α on a sufficiently small L-neighborhood N of β and for τ on a sufficiently small interval $[0, e]$. For $0 \leqq \tau \leqq e$, $\Delta(\eta, \beta)$ is thus a proper J-deformation of the L-neighborhood N of β. Hence β is homotopically ordinary and the proof of the theorem is complete.

7. The Riemannian manifold Σ

We turn now to a Riemannian manifold Σ with a metric defined by a positive definite quadratic form

$$ds^2 = g_{ij}(x)dx^i dx^j \qquad (i, j = 1, 2, \cdots, r).$$

Cf. Morse [1], p. 107. We suppose the functions $g_{ij}(x)$ are of class C'' in terms of the local coordinates (x) and that the transformations $z^i = z^i(x)$ from local coordinates (x) to local coordinates (z) are of class C'''. The manifold Σ will be assumed compact.

Given two points p and q on Σ there exists a path of least length joining p to q. This least length will be termed the distance $\delta(p, q)$. Given two rectifiable sensed curves α and β on Σ the Fréchet distance $\alpha\beta_\Sigma$ between α and β on Σ will be defined in the usual way using $\delta(p, q)$. The length of a curve α on Σ will be denoted by $\mathfrak{L}(\alpha)$ using a script $\mathfrak{L}$. The space of curves on Σ with the distances $\alpha\beta_\Sigma$ will be denoted by script $\mathfrak{M}$, and the space of curves on Σ with the distance

$$|\alpha\beta_\Sigma| = \alpha\beta_\Sigma + |\mathfrak{L}(\alpha) - \mathfrak{L}(\beta)|$$

will be denoted by $\mathfrak{L}$.

Let S be a convex region on a Cartesian n-space in which the variables (x) are admissible coordinates for Σ. A curve α on S and the image on Σ will be denoted by the same symbol. Let α and β be curves on S and Σ and let $\alpha\beta_S$ denote the Fréchet distance on S previously denoted by $\alpha\beta$. It is clear that for a fixed β, $\alpha\beta_S$ tends to 0 if and only if $\alpha\beta_\Sigma$ tends to 0.

If $\alpha\beta_\Sigma$ tends to 0 and $\mathfrak{L}(\alpha)$ tends to $\mathfrak{L}(\beta)$ we say that α *converges in $\mathfrak{L}$-length* to β. If α is on S the function $\mathfrak{L}(\alpha)$ is represented by an integral of the form

$$\int_{t_0}^{t_1} f(x, \dot{x})\, dt$$

with $f(x, r)$ satisfying all the conditions imposed on $f(x, r)$ in Morse and Ewing [1]. Hence for β on S, α converges in $\mathfrak{L}$-length to β if and only if α converges in L-length to β. Here L represents ordinary length on S.

An immediate technical problem is that of deducing various properties of an integral J along a curve of Σ from properties of J along subarcs, each in an admissible coordinate system. The problem of upper-reducibility of J is such a problem. The principal difficulty will be met with the aid of the following lemmas.

Let

$$v^i = \varphi^i(u) \qquad (i = 1, \cdots, n) \tag{7.1}$$

be an admissible transformation of coordinates neighboring a point $(u) = (u_0)$ with image $(v) = (v_0)$. Suppose $\varphi^i(u)$ has the form

$$\varphi^i(u) \equiv v_0^i + a_j^i(u^j - u_0^j) + \eta^i(u) \qquad a_j^i \equiv \frac{\partial \varphi^i}{\partial u_j}_{(u)=(u_0)}$$

The remainder $\eta^i(u)$ is of class C''' neighboring (u_0) since $\varphi^i(u)$ is of class C'''. In the space (u) let the solid n-sphere with radius a and center at (u_0) be denoted by σ_a. Our first lemma is as follows.

LEMMA 7.1. *Corresponding to the transformation T given by* (7.1) *there exists an admissible transformation $T_1 : v^i = \psi^i(u)$, of a neighbohrood $\sigma_{3\rho}$ of (u_0) such that*

$$\begin{aligned} \psi^i(u) &\equiv v_0^i + a_j^i(u^j - u_0^j) && [(u) \text{ on } \sigma_\rho] \\ \psi^i(u) &\equiv \varphi^i(u) && [(u) \text{ on } \sigma_{3\rho} - \sigma_{2\rho}] \end{aligned} \tag{7.2}$$

for a sufficiently small ρ.

To establish this lemma let $h(t)$ be a function of t of class C''' for t on the interval $0 \leqq t \leqq 1$ with

$$\begin{aligned} h(t) &\equiv 0 && (0 \leqq t \leqq 1/3) \\ h(t) &\equiv 1 && (2/3 \leqq t \leqq 1). \end{aligned}$$

To continue we simplify the proof by supposing that $(u_0) = (v_0) = (0)$. No generality is lost thereby. Corresponding to a positive constant ρ as yet undetermined consider the transformation

$$v^i = \psi^i(\rho, u) \equiv a_j^i u^j + h\left(\frac{|u|}{3\rho}\right)\eta^i(u) \qquad (i = 1, \cdots, n). \tag{7.3}$$

We shall show that this transformation satisfies the lemma if ρ is sufficiently small. Such a transformation satisfies (7.2) formally in the case $(u_0) = (v_0) = (0)$. It remains to show that for a suitable choice of ρ the transformation has a non-vanishing Jacobian and is one to one for (u) on $\sigma_{3\rho}$.

---

[6] Recall that the Jacobian $|a_j^1|$ does not vanish at $(u) = (u_0)$ for an admissible transformation.

To that end we make the substitution

$$v^i = \rho z^i \qquad u^i = \rho x^i \qquad (i = 1, \cdots, n) \tag{7.4}$$

in (7.3). We note that

$$\eta^i(\rho, x) = \rho^2 \zeta(\rho, x)$$

where ζ is of class C' in its arguments for $|x|$ bounded and ρ neighboring 0. In terms of the variables (z) and (x), (7.3) takes the form

$$z^i = a^i_j x^j + \rho h\left(\frac{|x|}{3}\right)\zeta(\rho, x). \tag{7.5}$$

For $\rho = 0$, (7.5) reduces to a non-singular collineation. If we restrict $|x|$ by the condition $|x| \leqq 3$, then for ρ sufficiently small the relation (7.5) is one to one and possesses a non-vanishing Jacobian. Hence for such a $\rho \neq 0$, (7.3) defines a one to one[7] transformation with a non-vanishing Jacobian. The proof of the lemma is complete.

A region S over which a system of admissible coordinates (u) of Σ range will be used to designate this system. We shall make use of the following lemma.

LEMMA 7.2. *Let S_1 and S_2 be two admissible coordinate systems such that the points of Σ represented by both S_1 and S_2 include a neighborhood of a point p. There then exists an admissible coordinate system S which represents the same points of Σ as does S_2 and which is such that for some neighborhood of p the transformation from the coordinates of S_1 to those of S is the identity.*

Let (u) and (v) be respectively the coordinates of S_2 and S_1 and suppose that (u_0) and (v_0) correspond to p. Suppose further that the transformation T given by (7.1) represents the relation between S_2 and S_1 neighboring (u_0) and (v_0), mapping a neighborhood of (u_0) on S_2 onto a neighborhood of (v_0) on S_1. Making use of the transformation T_1 of Lemma 7.1 set

$$T = T_1 T_2 \qquad [(u) \text{ on } \sigma_{3\rho}]$$

thereby defining T_2. We see that T_2 is the identity when (u) is on $\sigma_{3\rho} - \sigma_{2\rho}$ and so can be extended as the identity over $S_2 - \sigma_{3\rho}$. Let S_3 be the image of S_2 under T_2 so extended, and let a point of S_3 represent the same point of Σ as does its image on S_2.

Then T_1 gives the transformation from the coordinates, say (u), of S_3 to those of S_1, at least neighboring (u_0). But on a sufficiently small neighborhood of (u_0), T_1 is affine, and as such can be extended over the whole of S_3. Let S be the image of S_3 under this affine transformation and let a point of S represent the same point of Σ as does its image on S_3. Then the point with coordinates (v_0) on S corresponds to the point (v_0) on S_1, and the transformation from a point of S neighboring (v_0) to a point of S_1 neighboring (v_0) and representing the same point of Σ is the identity.

[7] This is not an ordinary inference from a non-vanishing Jacobian, but rather an inference in the large over all of $\sigma_{3\rho}$. It follows from the fact that for $\rho = 0$ the transformation is one to one over all of $\sigma_{3\rho}$.

8. The integral J on Σ

With each coordinate system (x) we have given a function $f(x, r)$ of (x) and a contravariant vector (r), with f invariant with respect to admissible changes of coordinates. Let α be a rectifiable curve given as the continuous image $p(t)$ on Σ of a t-interval $t_0 \leqq t \leqq t_1$ and with an absolutely continuous representation $x^i(t)$, $i = 1, \cdots, n$, in each coordinate system (x) in which α enters. In each such system set $f(x, \dot{x}) = \varphi(t)$. The value $\varphi(t)$ is independent of the coordinate system used to define $\varphi(t)$. We set

$$J(\alpha) = \int_{t_0}^{t_1} \varphi(t)\, dt.$$

That this integral exists as a Lebesgue integral, and is independent of the representations of α used to define it, follows from our assumption that $f(x, r)$ is of class C'' in (x) and (r) for $r \neq 0$, and that $f(x, r)$ is homogeneous in r in the sense of §5. These conditions will be assumed henceforth without explicit mention.

We also assume that $f(x, r)$ is convex in (r) in each coordinate system. That this is an invariant condition is seen from the fact that it is equivalent to the condition that the invariant Weierstrass E-function $E(x, r, \sigma)$ be non-negative. Here

$$E(x, r, \sigma) \equiv f(x, \sigma) - \sigma^i f_{r^i}(x, r) \qquad [(r) \neq (0)]$$

where (r) and (σ) are contravariant vectors.

The integrand $f(x, r)$ will be said to be *positive semi-normal* at a point $(x) = (c)$ if there exists a covariant vector (b) defined at (c) such that

$$f(c, r) > b_i r^i \qquad [(r) \neq 0] \tag{8.1}$$

for all non-null contravariant vectors (r) defined at (c).

The integral $J(\alpha)$ on Σ will be said to satisfy the *condition of Hahn* if the $\mathfrak{L}$-lengths of curves α on Σ on which $J(\alpha) \leqq 0$ are bounded.

In deriving the consequences of these conditions we shall have occasion to use the reduced μ-parameter of Morse, $0 \leqq \mu \leqq 1$, (to which we have referred earlier), along curves α. In defining this parameter one uses the metric of Σ defined by the distance $\delta(p, q)$. With this understood let α and β be curves on Σ. A subarc α' of α will be said to *correspond* to that subarc β' of β on which the μ-parameter ranges through the same values. As $\alpha\beta_\Sigma$ tends to 0, $\alpha'\beta'_\Sigma$ likewise tends to 0.

Lower semi-continuity of $J(\alpha)$. Suppose for a fixed β that $\alpha\beta_\Sigma$ tends to 0. Let β be divided into m successive arcs $\beta^{(1)}, \cdots, \beta^{(m)}$ such that $\beta^{(i)}$ lies in a coordinate system S_i. Let $\alpha^{(i)}$ be the subarc of α "corresponding" to $\beta^{(i)}$. As $\alpha\beta_\Sigma$ tends to 0, $\alpha^{(i)}\beta^{(i)}_{S_i}$ tends to 0 for each i, and conditions sufficient for the lower semi-continuity can be read off from the corresponding conditions in coordinate systems S.

In particular $J(\alpha)$ will be lower semi-continuous on any subset $\mathfrak{L}^*$ of $\mathfrak{L}$ provided f is convex in (r).

To show that the conditions sufficient that $\mathfrak{L}(\alpha)$ be bounded with $J(\alpha)$ are nominally the same as in a particular coordinate system we must revise the definition of a "pseudo-limiting" curve used in Morse and Ewing [1] in so far as that definition depends on the use of one coordinate system.

Without loss of generality we can suppose that each arc of unit $\mathfrak{L}$-length on Σ lies entirely in one coordinate system. This is a consequence of the fact that Σ can be covered by a finite number of such systems while an arc of unit $\mathfrak{L}$-length can be made arbitrarily small relative to distances in the coordinate systems by multiplying the form giving ds by a suitable positive constant.

A sequence of point representation $p_m(t)$ of curves on Σ will be said to *converge uniformly* to a representation $p(t)$, $t_0 \leqq t \leqq t_1$, if $p_m(t)$ converges to $p(t)$ for each t, uniformly with respect to t on $[t_0, t_1]$. We term $p(t)$ *absolutely continuous* if the functions $x^i(t)$ representing $p(t)$ in each coordinate system into which $p(t)$ enters are absolutely continuous.

We begin with the following lemma.

LEMMA 8.1. *Let $p_m(t)$, $m = 1, 2, \cdots$, be a sequence of arcs on Σ on each of which t is the $\mathfrak{L}$-length with $t_0 \leqq t \leqq t_0 + 1$. There then exists a subsequence $q_\mu(t)$, $\mu = 1, 2, \cdots$, of the sequence $p_m(t)$ which converges uniformly to an absolutely continuous point function $p(t)$.*

The arcs $p_m(t)$ have at least one cluster arc β to which a subsequence ω converges in the sense of Fréchet. The arc β lies in some coordinate system S since $\mathfrak{L}(\beta) \leqq 1$. In S let $x_\nu^i(t)$, $t_0 \leqq t \leqq t_0 + 1$, $\nu = 1, 2, \cdots$, be representations of those arcs of ω which lie in S. In accordance with Ascoli's theorem there exists a subsequence ω_1 of the sequence x_ν^i for which the corresponding functions x_ν^i converge uniformly on $[t_0, t_0 + 1]$ to absolutely continuous functions $X^i(t)$. Upon taking $p(t)$ as the point representation of the curve $X^i(t)$, $i = 1, \cdots, n$ the lemma follows.

A pseudo-limiting curve. Let $p_m(t)$, $m = 1, 2, \cdots$, be a sequence σ_0 of representations of arcs on Σ where t is the $\mathfrak{L}$-length, $0 \leqq t \leqq t_m$, and t_m becomes infinite with m. The sequence σ_0 will be said to have a *pseudo-limit* $p(t)$, if $p(t)$ is defined and absolutely continuous for $0 \leqq t < \infty$ and if $p_m(t)$ converges uniformly to $p(t)$ on each finite interval for t. It is naturally not required or expected that t be the $\mathfrak{L}$-length along $p(t)$.

LEMMA 8.2. *At least one subsequence of σ_0 possesses a pseudo-limit $P(t)$.*

In accordance with Lemma 8.1 there exists a subsequence σ_1 of σ_0 for which the corresponding subsequence of point functions $p_m(t)$ converges uniformly for $0 \leqq t \leqq 1$ to an absolutely continuous point function $P_1(t)$. We proceed inductively assuming that σ_{m-1} is a well defined subsequence of σ_{m-2}, $m = 3, 4, \cdots$. Using Lemma 8.1 again we infer the existence of a subsequence of σ_{m-1} such that the corresponding subsequence of functions $p_m(t)$ converges uniformly for $m - 1 \leqq t \leqq m$ to an absolutely continuous point function $P_m(t)$. We define $P(t)$ for $0 \leqq t < \infty$ by setting

$$P(t) \equiv P_m(t) \qquad (m - 1 \leqq t < m)$$
$$(m = 1, 2, \cdots).$$

Let σ be a subsequence $q_m(t)$ of σ_0 in which $q_m(t)$ is the m^{th} element of σ_m. It is clear that the sequence

$$q_m(t),\ q_{m+1}(t),\ \cdots$$

is a subsequence of σ_0 and that as m becomes infinite $q_m(t)$ converges uniformly to $P(t)$ on each finite interval for t.

Bounded $\mathfrak{M}$-compactness of $J(\alpha)$. We say that $J(\alpha)$ is boundedly $\mathfrak{M}$-compact if for each constant c the set J^c is an $\mathfrak{M}$-compact subset of $\mathfrak{M}$. By virtue of the lower semi-continuity of $J(\alpha)$ on sets $\mathfrak{L}^\kappa$, and with the use of the notion of pseudo-limiting curve one can prove the following theorem.

THEOREM 8.1. *If f is convex in (r), positive semi-normal in each coordinate system, and if J satisfies the condition of Hahn on Σ, then $J(\alpha)$ is boundedly $\mathfrak{M}$-compact and $\mathfrak{L}$ is bounded with J.*

One proves that $\mathfrak{L}$ is bounded with J as in the case of a single coordinate system. See Morse and Ewing [1], Theorems 3.3 and 3.2. The bounded $\mathfrak{M}$-compactness of J then follows from the $\mathfrak{M}$-compactness of $\mathfrak{L}^\kappa$ for each κ and the lower semi-continuity of $J(\alpha)$ on $\mathfrak{L}^\kappa$.

The proof of the upper-reducibility of $J(\alpha)$ involves new difficulties. The deformations $D(m, \beta)$ as defined in a coordinate system S, depend upon the notion of straightness in the system S. In deforming a curve α under $D(m, \beta)$ each point (x) on α is deformed along a straight line to its final destination. The difficulty is met with the aid of Lemma 7.2.

To prove the upper-reducibility of $J(\alpha)$ at a curve β we break β up into a sequence of arcs β_i, $i = 1, \cdots, \nu$, such that β_i lies entirely in some coordinate system $S^{(i)}$. Let p_i be the final end point of β_i. Then p_i lies both in $S^{(i)}$ and $S^{(i+1)}$. We apply Lemma 7.2 successively to $p_1, p_2, \cdots, p_{\nu-1}$. We are able thereby to affirm the following. There exists a sequence

$$S_1, S_2, S_3, \cdots, S_\nu \qquad [S_1 = S^{(1)}]$$

of coordinate systems such that S_i contains β_i and such that for a suitably chosen spherical neighborhood N_i of p_i, $i = 2, \cdots, \nu - 1$, the transformation on N_i from the coordinates of S_i to those of S_{i+1} is the identity.

We can now define a deformation $\theta(\alpha, \tau)$ of curves α on an $\mathfrak{M}$-neighborhood of β. Let α_i be the arc of α "corresponding" to β_i. We suppose $\alpha\beta_\Sigma$ so small that α_i lies in S_i, $i = 1, \cdots, \nu$, and that the final end point q_i of α_i, $i = 2, \cdots, \nu - 1$ lies on N_i. We subject α_i to the deformation $D(m, \beta_i)$ set up in S_i. The deformations $D(m, \beta_i)$ and $D(m, \beta_{i+1})$ replace q_i by the same point at the time τ. Hence these deformations combine to define a deformation $D(m, \beta)$ of α in which α is replaced at the time τ by $\theta(\alpha, \tau)$.

With $\theta(\alpha, \tau)$ so defined Lemma 5.1 holds formally, with L^κ replaced by $\mathfrak{L}^\kappa$, M by $\mathfrak{M}$, and $\varphi(\alpha, \tau)$ by $\theta(\alpha, \tau)$. We continue, except for notation, exactly as in the proof of Theorem 5.1. We then obtain the following theorem.

THEOREM 8.2. *If $\mathfrak{L}$ is bounded with J then J is upper-reducible at each rectifiable curve β on Σ.*

Conditions sufficient that $\mathfrak{L}$ be bounded with J are given in Theorem 8.1.

9. The existence of homotopic critical points of J

The principal conditions on a function $J(\alpha)$ defined on a metric space $\mathfrak{M}$ in order that the critical point theory apply are that each set J^c be compact and J be upper-reducible. We have seen that both these conditions are satisfied if for $(r) \neq 0$, f is of class C'', if f is convex in (r) and positive semi-normal in each coordinate system, and if J satisfies the condition of Hahn. We can now give simple conditions implying homotopic critical points.

We shall term the least upper bound of J on an arbitrary subset E of $\mathfrak{M}$ the *J-height* of E.

Our chains and cycles shall be defined on $\mathfrak{L}$ using $\mathfrak{L}$-continuity. They shall be finite singular chains and cycles, taken mod 2. See Morse [1], p. 146. The extension to the case of a field of coefficients is immediate if desired. We shall admit relative k-cycles u in which the modulus is always a set J^c in which c is less than the J-height of u.

The principal theorem on the existence of homotopic critical points is as follows. We assume that each set J^c is $\mathfrak{M}$-compact.[8]

THEOREM 9.1. *Let K be a homology class of k-cycles mod J^a, non-bounding mod J^a on $\mathfrak{L}$ and let c be the greatest lower bound of J-heights of cycles of K. If $c > a$, and if J is upper-reducible at each point of J^c there exists a homotopic critical point at which $J = c$.*

The theorem is a consequence of Corollary 3.3 of Morse [4].

If each set J^c is $\mathfrak{M}$-compact then J is bounded below on $\mathfrak{M}$. For a sequence of points α for which J became negatively infinite could converge to no point β of $\mathfrak{M}$. For $J(\beta)$ would necessarily be less than every constant c.

If the constant a of Theorem 9.1 is less than every value of J the homology class K of Theorem 9.1 becomes a class K of absolute cycles corresponding to which the theorem affirms the existence of a homotopic critical point at which $J = c$. If K is the homology class of 0-cycles this specialization gives a critical point affording an absolute minimum to J on Σ. Theorem 9.1 may be employed in a variety of ways to obtain and classify critical points. Homotopic critical points lead to "extremals" as will be shown in §10.

10. The existence of extremals

Let β be a curve of $\mathfrak{M}$. We shall term β an *extremal* if each subarc β^* of β lying in a coordinate system S is an extremal of J in the coordinate system S.

We wish to use the definition of a homotopic critical point, replacing the L-neighborhood in this definition by a J-neighborhood. This is permissible if each $\mathfrak{L}$-neighborhood of β contains a J-neighborhood of β. To give conditions for this we must extend certain definitions to Σ.

If (x) is a point on β in a coordinate system S, a pair (x, r) in which (r) is a non-null contravariant vector at (x) will be termed an *element tangent* to β at (x) if (x, r) is an element tangent to β at (x) in the ordinary sense in S. The

---

[8] Sufficient conditions that J^c be $\mathfrak{M}$-compact are given in Theorem 8.1.

phrase *almost all elements tangent* to β is used for a set of elements tangent to β at almost all points s of β, where s is the arc length on β.

The condition that f be *strongly convex* at an element (x, ρ), where ρ is a non-null contravariant vector, at a point (x) in a coordinate system S, is that the invariant E-function $E(x, \rho, r)$ be positive for the given (x, ρ), except when $r = k\rho$ with $k \geqq 0$.

With these definitions we can reaffirm Theorem 6.0 for Σ with the new interpretation of its terms. Thus, if f is strongly convex at almost all elements tangent to β and is positive semi-normal at all points of β, then α converges in $\mathfrak{L}$-length to β, if α converges in J-length to β.

We are ready for the homotopy theorem.

THEOREM 10.1. *If convergence in J-length to a curve β implies convergence in $\mathfrak{L}$-length to β then β is an extremal whenever homotopically critical.*

We begin with a proof of statement (a).

(a). *When convergence in J-length to β implies convergence in $\mathfrak{L}$-length to β, then convergence in J-length to a subarc β_1 of β implies convergence in $\mathfrak{L}$-length and L-length to β_1.*

Suppose β is a sequence of the three arcs β_0, β_1, β_2, admitting the possibility that β_0 or β_2 reduce to points. Let p_1 and q_1 be the end points of β_1, and suppose p_1 and q_1 be in convex coordinate systems S and T. Let α_1 be an arc so near β_1 that its end points p and q be on S and T respectively. Let $\overline{p_1p}$ and $\overline{qq_1}$ be line segments on S and T respectively joining p_1 to p and q to q_1. Let α_1 be extended by forming the curve

$$\alpha = \beta_0 \overline{p_1p} \alpha_1 \overline{qq_1} \beta_2 .$$

When α_1 converges in J-length to β_1 it is clear that α converges in J-length to β, so that α then converges in $\mathfrak{L}$-length to β, and finally α_1 converges in $\mathfrak{L}$-length to β_1.

Since convergence in $\mathfrak{L}$-length and L-length are equivalent the proof of (a) is complete.

Theorem 10.1 will be established in its equivalent form: if the first variation $I(\eta)$, formed for a subarc β^* of β in some coordinate system S, is negative for some admissible η, and if the hypothesis of (a) holds then β is homotopically ordinary. By virtue of (a), convergence in J-length to β^* implies convergence in L-length to β^*. In accordance with the results of §6 there accordingly exists a proper J-deformation $\Delta(\eta, \beta)$ of an L-neighborhood N^* of β^*.

Let α be a curve on so small an $\mathfrak{L}$-neighborhood N of β that the subarc α^* of α "corresponding" to β^* is on N^*. The point deformation used to define $\Delta(\eta, \beta^*)$ will not move the end points of α^* and so can be regarded as defining a deformation D of α in which α^* alone varies. This deformation D is a proper J-deformation of the $\mathfrak{L}$-neighborhood N of β. Under the hypotheses of the theorem every $\mathfrak{L}$-neighborhood of β contains a J-neighborhood of β. That β is homotopically ordinary now follows from the definition involved.

11. Résumé

We shall not attempt in any sense to summarize our results under the weakest hypotheses under which they are proved. A broad résumé will however be useful.

In order that J^c be compact for each constant c and J be upper-reducible it is sufficient that f be of class C' for $(r) \neq (0)$, homogeneous in (r) in the usual sense, convex in (r), and positive semi-normal in each coordinate system, and that J satisfy the condition of Hahn. Then the existence of homology classes of non-bounding k-cycles or relative k-cycles implies the existence of homotopic critical points. (See Theorem 9.1.)

If β is a homotopic critical point, the conditions sufficient that β be an "extremal" are somewhat stronger but the additional condition need be satisfied only for points (x) near β. The additional sufficient conditions are that f be strongly convex at almost all elements tangent to β and be of class C'' for (x) neighboring β and for $(r) \neq (0)$.

If f is positive for every $(r) \neq (0)$ the condition of Hahn is fulfilled. If f is positive and positive regular in the classical sense, then all of the preceding conditions are satisfied. Thus the classical theory in the large becomes a special case of the preceding, at least in its analytical as distinguished from its topological foundations.

INSTITUTE FOR ADVANCED STUDY,
UNIVERSITY OF MISSOURI.

BIBLIOGRAPHY

McSHANE, E. J.

1. *Semicontinuity of integrals in the calculus of variations*, Duke Mathematical Journal, 2 (1936), 597–616.
2. *Curve-space topologies associated with variational problems*, Annali della R. Scuola Normale Superiore di Pisa, Serie II, 9 (1940), 45–60.

MORSE, M.

1. *The calculus of variations in the large*, American Mathematical Society Colloquium Publications XVIII (1934).
2. *A special parameterization of curves*, Bulletin of the American Mathematical Society 42 (1936), 915–22.
3. *Functional topology and abstract variational theory*, Mémorial des Sciences Mathématiques, XCII, Gauthier-Villars, Paris (1939).
4. *Functional topology*, Bulletin of the American Mathematical Society, 49 (1943), pp. 144–149.

MORSE, M. AND EWING, G.

1. *The variational theory in the large including the non-regular case*—first paper, Annals of Mathematics (1943).

MORSE, M. AND TOMPKINS, C.

1. *The existence of minimal surfaces of general critical types*, Annals of Mathematics 40 (1939), 443–72.

ANNALS OF MATHEMATICS
Vol. 44, No. 3, July, 1943

ANALYTIC SOLUTIONS OF NON-LINEAR DIFFERENCE EQUATIONS

BY WALTER STRODT

(Received February 17, 1943)

Introduction

This paper treats a broad class of ordinary difference equations, both linear and non-linear, with arbitrarily many not necessarily commensurate spans. The coefficients in the equation are rational functions, or certain more general functions of x. Solutions are obtained which are analytic in a half-plane and which satisfy a condition of restricted growth at infinity. For each equation the number of such solutions obtained is precisely the degree of the equation in the unknown, and it is proved that there are no solutions, other than these, analytic in such a half-plane and satisfying such a condition of restricted growth. The solutions are found by a new technique constructed from a concept of approximating q-difference equations, from a procedure analogous to the calculus of limits as applied to algebraic functions, from the classic compactness theorem for bounded families of analytic functions, and from a theory of special functions ("almost constant" functions) which are generalizations of rational functions bounded at infinity.

To sketch the method more precisely, let us consider its application to the very special case of a difference equation of the form

$$(\alpha) \qquad y(x + \omega_1)y(x + \omega_2) + y(x + \omega_3)y(x + \omega_4) = r(x),$$

where ω_1, ω_2, ω_3, ω_4 are non-negative numbers, and $r(x)$ is a rational function whose limit at infinity is finite and different from zero.

We shall, without essential loss of generality in this case, assume that $\omega_1 = 0$.

Letting b be any positive number greater than every ω_s, and defining $q_s = 1 - \omega_s b^{-1}$, $(s = 1, 2, 3, 4)$, we consider the "approximating q-difference equation"

$$(\beta) \qquad y(q_1x + \omega_1)y(q_2x + \omega_2) + y(q_3x + \omega_3)y(q_4x + \omega_4) = r(x).$$

When b tends to infinity, this equation tends formally to the given difference equation, since every q_s tends to unity.

Now if for every sufficiently large value of b there is an analytic solution $y(x, b)$ of (β), and if these solutions have a region of analyticity $\mathfrak{D}$ in common, and are bounded in $\mathfrak{D}$ by a bound M independent of b, then the compactness theorem[1] for bounded families of analytic functions shows that there exists a sequence $\{b_k\,;\, k = 1, 2, \cdots\}$ of positive numbers tending to infinity for which the sequence $\{y(x, b_k);\, k = 1, 2, \cdots\}$ tends to a limit function, uniformly in

[1] Montel, "Leçons sur les familles normales," Paris 1927, section 10.

every closed bounded subset of $\mathfrak{D}$. If $\mathfrak{D}$ is sufficiently extensive, such a limit function evidently is an analytic solution of (α).

We consider the question of finding such $y(x, b)$, $\mathfrak{D}$, and M.

The point $x = b$ plays an important role in this question, since if $y(x) = \sum_{j=0}^{\infty} c_j(x - b)^j$, then $y(q_s x + \omega_s) = \sum_{j=0}^{\infty} c_j q_s^j(x - b)^j$, and (β) becomes

$$(\gamma) \qquad \begin{aligned}&[\textstyle\sum_{j=0}^{\infty} c_j q_1^j(x - b)^j][\sum_{j=0}^{\infty} c_j q_2^j(x - b)^j] \\ &\qquad + [\textstyle\sum_{j=0}^{\infty} c_j q_3^j(x - b)^j][\sum_{j=0}^{\infty} c_j q_4^j(x - b)^j] = r(x).\end{aligned}$$

It is the simplicity of this equation which makes the introduction of equation (β) advantageous.

Letting $r(x) = \sum_{j=0}^{\infty} r_j(x - b)^j$, we see that equation (γ) determines the c_j recursively as follows:

$$2c_0^2 = r_0$$

$$\begin{aligned}c_0c_j(q_1^j + q_2^j + q_3^j + q_4^j) = r_j& \\ &-c_1c_{j-1}(q_1q_2^{j-1} + q_2q_1^{j-1} + q_3q_4^{j-1} + q_4q_3^{j-1}) \\ &-c_2c_{j-2}(q_1^2q_2^{j-2} + q_2^2q_1^{j-2} + q_3^2q_4^{j-2} + q_4^2q_3^{j-2}) \\ &- \cdots, \qquad\qquad (j = 1, 2, \cdots).\end{aligned}$$

Since r_0 approaches a finite non-zero limit as b becomes infinite, there exist positive numbers M_1, M_2 independent of b such that $M_1 < | c_0 | < M_2$ for all large values of b. Moreover, $0 < q_s \leqq 1$, $(s = 1, 2, 3, 4)$, and $q_1 = 1$. Hence, if numbers C_j, $(j = 1, 2, \cdots)$, are defined recursively by the equations

$$M_1C_j = | r_j | + 4C_1C_{j-1} + 4C_2C_{j-2} + \cdots,$$

the inequalities $| c_j | \leqq C_j$ will be valid, $(j = 1, 2, \cdots)$. Now if $C(x, b) = \sum_{j=1}^{\infty} C_j(x - b)^j$, and $R(x, b) = \sum_{j=1}^{\infty} | r_j | (x - b)^j$, $C(x, b)$ will satisfy the algebraic equation

$$M_1C(x, b) = R(x, b) + 2C^2(x, b).$$

Thus $C(x, b)$ is analytic wherever $R(x, b)$ is analytic, except possibly at values of x for which $M_1^2 = 8R(x, b)$. If D_0 is a number exceeding the modulus of every pole of $r(x)$, $R(x, b)$ is analytic when $| x - b | < b - D_0$. Moreover, for every positive ϵ there is a number $D(\epsilon)$ independent of b such that $R(x, b)$ is analytic and satisfies the inequality $| R(x, b) | < \epsilon$ throughout the region $| x - b | < b - D(\epsilon)$. (This ϵ-property of $r(x)$ is the characteristic property of the class of functions designated in this paper as "almost constant" functions[2]. It is possessed by all rational functions which are bounded at infinity[3], and by many other functions, including for example the solutions of (α) which we shall obtain.)

[2] Definition 2, below.

[3] By Lemma VII of the appendix.

If now ϵ_0 is any positive number smaller than $(M_1^2)/8$, the equation $M_1^2 = 8R(x, b)$ is false at every point of the region $|x - b| < b - D(\epsilon_0)$, and consequently $C(x, b)$ is analytic in the region $|x - b| < b - D(\epsilon_0)$. Since $|c_j| \leqq C_j$, $(j = 1, 2, \cdots)$, it follows that $\sum_{j=0}^{\infty} c_j(x - b)^j$ is analytic in the region $|x - b| < b - D(\epsilon_0)$. By choosing ϵ_0 sufficiently small we can make $R(x, b)$ arbitrarily small in $|x - b| < b - D(\epsilon_0)$, and hence make $C(x, b)$ arbitrarily near a root X of the equation $M_1X = 2X^2$. Thus, if ϵ_0 is fixed as a sufficiently small number, there will be a number M_0 independent of b such that $|C(x, b)| < M_0$ throughout $|x - b| < b - D(\epsilon_0)$. This implies that $|\sum_{j=1}^{\infty} c_j(x - b)^j| < M_0$ throughout $|x - b| < b - D(\epsilon_0)$. Thus we shall obtain the desired $y(x, b)$, $\mathfrak{D}$, and M by taking for $y(x, b)$ the series $\sum_{j=0}^{\infty} c_j(x - b)^j$, taking for $\mathfrak{D}$ any bounded region which with its boundary is included in the half-plane $\Re(x) > D(\epsilon_0)$, and taking for M the number $M_0 + M_2$.

Hence at least one solution $y(x)$ of equation (α), analytic in $\mathfrak{D}$, is obtained as the limit of a sequence $\{y(x, b); b = b_1, b_2, \cdots\}$ of solutions of equations (β). This solution of (α) can be continued analytically throughout the half-plane $\Re(x) > D(\epsilon_0)$ by the use of an increasing sequence of regions $\mathfrak{D}$. It will be bounded in this half-plane by the bound M. Since there are two distinct choices for c_0, it is easy to prove that at *least* two distinct solutions of (α), analytic and bounded in a right half-plane, can be obtained in this way.

It can be shown that every solution of (α) which is analytic and bounded in a right half-plane can be expressed as the limit, as b become infinite through *all* sufficiently large positive values, of $y(x, b)$, where $y(x, b)$ is a solution of (β), analytic at $x = b$, and where the limit is uniform in every closed bounded subset of a certain right half-plane; from this it follows easily that there are at *most* two distinct solutions of (α), analytic and bounded in a right half-plane. For the proof that each bounded solution of (α) can be so expressed as a limit function, we note that if $y_0(x)$ is a solution of (α) then it is also a solution of the q-difference equation

$$(\delta) \qquad y(q_1x + \omega_1)y(q_2x + \omega_2) + y(q_3x + \omega_3)y(q_4x + \omega_4) = \rho(x),$$

where
$$\begin{aligned}\rho(x) = r(x) + [&y_0(q_1x + \omega_1)y_0(q_2x + \omega_2)\\ &- y_0(x + \omega_1)y_0(x + \omega_2)\\ &+ y_0(q_3x + \omega_3)y_0(q_4x + \omega_4)\\ &- y_0(x + \omega_3)y_0(x + \omega_4)].\end{aligned}$$

A study of the difference $f(q_sx + \omega_s) - f(x + \omega_s)$, $(s = 1, 2, 3, 4)$, for functions $f(x)$ bounded in a right half-plane shows that if $y_0(x)$ is bounded in a right half-plane, $\rho(x)$ is in a certain sense so nearly equal to $r(x)$ when b is large that one of the two solutions of (β), analytic at $x = b$, is near the solution $y_0(x)$ of (δ). The detailed proof of this part, even in the simple case under discussion, leans

heavily upon the lemmas of the appendix to this paper, where a systematic study of almost constant functions is made.

The proof of the general theorems demonstrated below follows closely the outline of the proof in this special case, except that a modification of the technique of dominant functions is employed; this modification consists in a transformation of the approximating q-difference equation by a substitution of the form $y(x) = x^{\sigma}z(x)$ before the use of dominant functions in the manner described above; the purpose of the modification is the securing of a sharper estimate for the radius of convergence of the solutions at $x = b$ of the approximating q-difference equation.

We now state the general theorems to be proved in this paper.

Statement of Theorems

THEOREM 1. *Given the difference equation of degree n,*

$$(1) \qquad \sum_{k=1}^{m} a_k(x) \prod_{s=1}^{s(k)} y(x + \omega_{ks}) = \varphi(x),$$

where

(2) *$\varphi(x)$ is a rational function, and therefore is asymptotically equivalent to cx^p for some non-zero complex number c and some integer p, (positive, negative, or zero),*

(3) *$a_k(x)$ is a rational function having a finite (perhaps zero) limit a_k at infinity, $(k = 1, 2, \cdots, m)$,*

(4) $0 \leqq \omega_{k1} \leqq \omega_{k2} \leqq \cdots \leqq \omega_{k,s(k)}$, $\qquad (k = 1, 2, \cdots, m)$,

(5) $s(k) \leqq n$, $\qquad (k = 1, 2, \cdots, m)$,

(6) *$\Re(a_k) \geqq 0$ whenever $s(k) = n$,*

(7) *$s(k) = n$ for all k if $p \leqq 0$,*

(8) *for at least one value of k, the relations $s(k) = n$, $\Re(a_k) > 0$, $\omega_{k1} = 0$ are all valid.*

Conclusion: For all sufficiently large positive constants D, M there are in the region $\Re(x) > D$ exactly n analytic solutions $y(x)$ of (1) *which satisfy the inequality* (9) $| y(x) | \leqq M | x |^{p/n}$.

Every solution $y(x)$ of (1) *which is analytic in a half-plane $\Re(x) > D > 0$ and there satisfies* (9) *can be expressed in the following form:*

$$(10) \qquad y(x) = \lim_{b \to \infty} y_b(x)$$

where $y_b(x)$ is a solution of the "approximating q-difference equation" (see *Definition* 1), *analytic at $x = b$, and where the limit is uniform in every closed bounded subset of some half-plane $\Re(x) > D' \geqq D$.*

DEFINITION 1. THE APPROXIMATING q-DIFFERENCE EQUATION. *Let b be a positive*[4] *number greater than every one of the numbers ω_{ks} appearing in equation* (1). *Let q_{ks} be defined by the equations*

---

[4] Throughout this paper it is always understood that b is positive.

$$(11) \qquad q_{ks} = 1 - \omega_{ks} b^{-1}, \qquad (k = 1, \cdots, m; s = 1, 2, \cdots, s(k)).$$

Then the functional equation

$$(12) \qquad \sum_{k=1}^{m} a_k(x) \prod_{s=1}^{s(k)} y(q_{ks} x + \omega_{ks}) = \varphi(x)$$

will be called the approximating q-difference equation for equation (1).

(This terminology is motivated by the fact demonstrated in Lemma 1 that (12) becomes a q-difference equation if the independent variable is taken as $x - b$ instead of x, and by the fact that when b is large q_{ks} is near 1, so that equation (12) is in a formal sense an approximation to equation (1). It may be seen from this that the "singular point" of (12), from the standpoint of the theory of q-difference equations, is the point $x = b$, which is precisely the point at which $y_b(x)$ is required to be analytic.)

The proof of Theorem 1 will be omitted, since in what follows a more general theorem will be demonstrated. The condition that $a_k(x)$, $(k = 1, \cdots, m)$, and $\varphi(x)x^{-p}$ be rational and bounded at infinity will be replaced by the less restrictive condition that these functions be "almost constant", as explained in Definition 2. (It is a consequence of Lemma VII of the appendix that every rational function which is bounded at infinity is almost constant.)

Definition 2. Almost Constant Functions. *Let f be either a function $f(x)$ which is analytic in a half-plane $\Re(x) > D > 0$, or a function $f(x, b)$ of the two variables x, b which for all sufficiently large values of b is analytic in the region $|x - b| < b - D$, for a positive number D independent of b. Then if b is sufficiently large, f can be expanded in a Taylor's series $\sum_{j=0}^{\infty} \beta_j (x - b)^j$, convergent for $|x - b| < b - D$. We shall say that f is almost constant if for every positive ϵ there is a positive number $D(\epsilon)$, independent of b, such that $|\sum_{j=0}^{\infty} |\beta_j| (x - b)^j - |\beta_0|| < \epsilon$ whenever $|x - b| < b - D(\epsilon)$.* (Note: In what follows, we shall sometimes deal with functions of the two variables x, b, of which only the first is displayed in the notation.)

We now state the more general theorem.

Theorem 2. *Given the difference equation* (1), *where*

(2′) *$\varphi(x)x^{-p}$ is almost constant, and the limit as b becomes infinite of $\varphi(b)b^{-p}$ is c, for some* real *number p and some non-zero complex number c,*

where

(3′) *$a_k(x)$ is almost constant, and the limit as b becomes infinite of $a_k(b)$ is a finite (perhaps zero) number a_k, $(k = 1, 2, \cdots, m)$,*

and where (4), (5), (6), (7), (8) *are valid.*

Conclusion: All the conclusions of Theorem 1 *are valid, with this new interpretation of equation* (1). *Moreover, if $y_0(x)$ is any solution of* (1) *which is analytic in a half-plane $\Re(x) > D > 0$ and there satisfies* (9), *then $y_0(x)x^{-p/n}$ is almost constant.*

Proof of Theorem 2

PART I. PRELIMINARY MODIFICATIONS OF THE APPROXIMATING q-DIFFERENCE EQUATION

LEMMA 1. *Let $f(x)$ be any function of x, analytic at $x = b$, and suppose that*

$$f(x) = \sum_{j=0}^{\infty} c_j(x-b)^j. \tag{13}$$

Then $f(q_{ks}x + \omega_{ks})$ is analytic at $x = b$, and

$$f(q_{ks}x + \omega_{ks}) = \sum_{j=0}^{\infty} c_j q_{ks}^j (x-b)^j, \qquad (k+1, \cdots, m;\ s = 1, \cdots, s(k)). \tag{14}$$

PROOF.

$$\begin{aligned} f(q_{ks}x + \omega_{ks}) &= \sum_{j=0}^{\infty} c_j(q_{ks}x + \omega_{ks} - b)^j \\ &= \sum_{j=0}^{\infty} c_j(q_{ks}x + b - bq_{ks} - b)^j = \sum_{j=0}^{\infty} c_j q_{ks}^j (x-b)^j. \end{aligned}$$

LEMMA 2. *The substitution*[5] $y(x) = x^{p/n}z(x)$ *transforms equation* (12) *into*

$$\sum_{k=1}^{m} \alpha_k(x) \prod_{s=1}^{s(k)} z(q_{ks}x + \omega_{ks}) = r(x), \tag{15}$$

where

$$\alpha_k(x) = a_k(x)x^{-p(n-s(k))/n} \prod_{s=1}^{s(k)} (q_{ks} + \omega_{ks}x^{-1})^{p/n}, \qquad (k = 1, \cdots, m), \tag{16}$$

and

$$r(x) = \varphi(x)x^{-p}. \tag{17}$$

LEMMA 3. *Let the definitions* $\alpha_k(x) = \sum_{j=0}^{\infty} \alpha_{kj}u^j$, $r(x) = \sum_{j=0}^{\infty} r_ju^j$, $z(x) = \zeta + \sum_{j=1}^{\infty} z_ju^j$ *be made, where*

$$u = x - b. \tag{18}$$

(The functions $\alpha_k(x)$, $r(x)$ are easily seen to be analytic at $x = b$, and hence so expressible, provided b is sufficiently large; ζ and the z_j on the other hand are to be determined.) *Then equation* (15) *becomes the q-difference equation*

$$\sum_{k=1}^{m}\left(\sum_{j=0}^{\infty} \alpha_{kj}u^j\right) \prod_{s=1}^{s(k)} \left(\zeta + \sum_{j=1}^{\infty} z_j q_{ks}^j u^j\right) - \sum_{j=0}^{\infty} r_j u^j = 0. \tag{19}$$

Proof by Lemma 1.

---

[5] Here and throughout the paper that branch of $x^{p/n}$ in $\Re(x) > 0$ is chosen which is positive when x is positive.

PART II. FORMAL SOLUTIONS OF THE q-DIFFERENCE EQUATION (19)

LEMMA 4. *If the coefficient of* u^j, $(j = 0, 1, 2, \cdots)$ *in the left-hand member of* (19) *is equated to zero, the equations obtained are*

$$\sum_{k=1}^{m} \alpha_{k0}\zeta^{s(k)} = r_0, \tag{20}$$

and

$$z_j \sum_{k=1}^{m} \alpha_{k0}\zeta^{s(k)-1}(q_{k1}^j + q_{k2}^j + \cdots + q_{k,s(k)}^j) = r_j - P_j(\zeta, z_1, z_2, \cdots, z_{j-1}, q_{ks}, \alpha_{i\vartheta}), \tag{21}$$

where P_j, $(j = 1, 2, \cdots)$ *is a polynomial, with positive integers for coefficients, in the indicated arguments,* (*with* $i, k = 1, 2, \cdots, m$; $s = 1, 2, \cdots, s(k)$; $\vartheta = 0, 1, \cdots, j$).

LEMMA 5.

$$\lim r_0 = c.$$

$$\lim \alpha_{k0} = 0, \text{ if } s(k) \neq n.$$

$$\lim \alpha_{k0} = a_k, \text{ if } s(k) = n. \qquad (\textit{All limits as } b \to \infty)$$

$$\Re\left(\sum_{s(k)=n} \sum_{s=1}^{n} a_k q_{ks}^j\right)$$

is bounded below by a positive number independent of j *and* b, $(j = 0, 1, 2, \cdots)$.

PROOF. The first assertion follows from (2′), and the fact that $r(x) = \varphi(x)x^{-p}$. Because of (7), the second assertion is vacuously true when $p \leqq 0$; when $p > 0$, it follows at once from (16). The third assertion follows from (16), since $q_{ks} \to 1$ when $b \to \infty$, and since (3′) states that $a_k(b) \to a_k$ when $b \to \infty$.

The fourth assertion follows from the fact that because of (8) there is a number l such that $s(l) = n$, $\Re(a_l) > 0$, and $\omega_{l1} = 0$. From this $q_{l1} = 1$, and therefore, in view of (6), we have

$$\Re\left(\sum_{s(k)=n} \sum_{s=1}^{n} a_k q_{ks}^j\right) \geqq \Re\left(\sum_{s=1}^{n} a_l q_{ls}^j\right) \geqq \Re(a_l),$$

and this lower bound is independent of j and b.

LEMMA 6. *When b is large, equation* (20) *has n solutions* $\zeta = \zeta^{(t)}$, $(t = 1, 2, \cdots, n)$ *one near each of the n distinct n^{th} roots* $\sigma_1, \sigma_2, \cdots, \sigma_n$ *of the number* c/I, *where* $I = \sum a_k$, *the summation being over all the values of k for which* $s(k) = n$. *There exist positive numbers*[6] M_1, M_2 *such that*

$$M_1 < |\zeta^{(t)}| < M_2, \qquad (t = 1, 2, \cdots, n), \tag{22}$$

for all sufficiently large values of b.

PROOF. By the last part of Lemma 5, $I \neq 0$.

[6] Here, and henceforth, the symbols M and D, possibly with subscripts, shall stand for positive numbers independent of b.

By Lemma 5, as $b \to \infty$, the initial coefficient in (20) tends to the non-zero number I; the constant term tends to the non-zero number c, and all other coefficients tend to zero.

Since the roots of a polynomial are continuous functions of the coefficients, Lemma 6 follows.

LEMMA 7. *For all sufficiently large values of b, the coefficient of z_j in (21) is in modulus bounded below by a positive number M_3 independent of j.*

PROOF. The coefficient in question is $\sum_{k=1}^{m} \sum_{s=1}^{s(k)} \alpha_{k0}\zeta^{s(k)-1} q_{ks}^{j}$. When b is large the terms in this summation for which $s(k) \neq n$ are small, because of (22) and Lemma 5. The remaining terms are near $\sum_{s(k)=n} \sum_{s=1}^{n} a_k(\sigma_t)^{n-1} q_{ks}^{j}$, (for some t), the modulus of which is at least $M_1^{n-1}\Re(\sum_{s(k)=n} \sum_{s=1}^{n} a_k q_{ks}^{j})$, and this by Lemma 5 is bounded below by a positive number independent of j.

LEMMA 8. *If b is large equation (19) is satisfied by exactly n distinct formal power series $\zeta^{(t)} + \sum_{j=1}^{\infty} z_j^{(t)} u^j$, $(t = 1, 2, \cdots, n)$.*

PROOF. By Lemmas 6 and 7, (since the coefficient of z_j in equation (21), being bounded from below, cannot be zero).

PART III. ANALYTICITY OF DOMINANT FUNCTIONS

LEMMA 9. *Let Z_j, $(j = 1, 2, \cdots)$ be defined by the recursive relations*

$$Z_j M_3 = |r_j| + P_j(M_2, Z_1, Z_2, \cdots, Z_{j-1}, 1, |\alpha_{i\vartheta}|). \tag{23}$$

Then $|z_j^{(t)}| \leqq Z_j$, $(t = 1, 2, \cdots, n; j = 1, 2, \cdots; i = 1, 2, \cdots, m; \vartheta = 0, 1, \cdots, j)$.

PROOF. This follows from (21), (22), Lemma 7, and the fact that the coefficients in the polynomial P_j are positive. (A formal proof can be given by induction on j.)

LEMMA 10. *Let $Z(u)$ be the formal power series $\sum_{j=1}^{\infty} Z_j u^j$. Then $Z(u)$ satisfies the following algebraic equation*

$$\begin{aligned} M_3 Z(u) &= R(u) - R(0) + \sum_{k=1}^{m} [A_k(u) - A_k(0)][M_2 + Z(u)]^{s(k)} \\ &\quad + \sum_{k=1}^{m} A_k(0)[(M_2 + Z(u))^{s(k)} - s(k)M_2^{s(k)-1}Z(u) - M_2^{s(k)}], \end{aligned} \tag{24}$$

where

$$R(u) = \sum_{0}^{\infty} |r_j| u^j, \tag{25}$$

and

$$A_k(u) = \sum_{0}^{\infty} |\alpha_{kj}| u^j, \qquad (k = 1, 2, \cdots, m). \tag{26}$$

Proof. From (19), (20), and (21) it is evident that

$$\sum_{j=1}^{\infty} u^j P_j(\zeta, z_1, z_2, \cdots, z_{j-1}, q_{ks}, \alpha_{i\vartheta}) = \sum_{k=1}^{m} \left(\sum_{j=0}^{\infty} \alpha_{kj} u^j\right) \prod_{s=1}^{s(k)} \left(\zeta + \sum_{j=1}^{\infty} z_j q_{ks}^j u^j\right)$$
$$- \sum_{j=1}^{\infty} u^j \sum_{k=1}^{m} \alpha_{k0} \zeta^{s(k)-1} z_j (q_{k1}^j + q_{k2}^j + \cdots + q_{k,s(k)}^j) - \sum_{k=1}^{m} \alpha_{k0} \zeta^{s(k)}.$$

Hence

$$\sum_{j=1}^{\infty} u^j P_j(M_2, Z_1, Z_2, \cdots, Z_{j-1}, 1, |\alpha_{i\vartheta}|)$$
$$= \sum_{k=1}^{m} \left(\sum_{j=0}^{\infty} |\alpha_{kj}| u^j\right) \prod_{s=1}^{s(k)} \left(M_2 + \sum_{j=1}^{\infty} Z_j u^j\right)$$
$$- \sum_{j=1}^{\infty} u^j \sum_{k=1}^{m} |\alpha_{k0}| M_2^{s(k)-1} Z_j s(k) - \sum_{k=1}^{m} |\alpha_{k0}| M_2^{s(k)}$$
$$= \sum_{k=1}^{m} A_k(u)(M_2 + Z(u))^{s(k)}$$
$$- \sum_{k=1}^{m} M_2^{s(k)-1} s(k) A_k(0) Z(u) - \sum_{k=1}^{m} A_k(0) M_2^{s(k)}$$
$$= \sum_{k=1}^{m} (A_k(u) - A_k(0))(M_2 + Z(u))^{s(k)}$$
$$+ \sum_{k=1}^{m} A_k(0)[(M_2 + Z(u))^{s(k)} - s(k) M_2^{s(k)-1} Z(u) - M_2^{s(k)}].$$

But (23) implies that

$$M_3 Z(u) = R(u) - R(0) + \sum_{j=1}^{\infty} P_j(M_2, Z_1, Z_2, \cdots, Z_{j-1}, 1, |\alpha_{i\vartheta}|) u^j.$$

Hence Lemma 10 is established.

Lemma 11. *For every positive ϵ there is a positive number $D(\epsilon)$ such that for all sufficiently large values of b*

$$(27) \qquad |R(u) - R(0)| < \epsilon \quad \textit{when} \quad |u| < b - D(\epsilon),$$

and

$$(28) \quad |A_k(u) - A_k(0)| < \epsilon \quad \textit{when} \quad |u| < b - D(\epsilon), \qquad (k = 1, 2, \cdots, m).$$

Proof. $r(x)$ is almost constant, by hypothesis. Since by hypothesis $a_k(x)$ is almost constant, it follows from Lemmas II, III, VI of the appendix that $\alpha_k(x)$ is almost constant.

Lemma 12. *For every positive δ there is a positive number $D_1(\delta)$ such that throughout the region $|u| < b - D_1(\delta)$, for all sufficiently large values of b, every coef-*

ficient of (24), *considered as an equation in* $Z(u)$, *is within* δ *of the corresponding coefficient in the equation*

$$(29) \qquad M_3 X = \sum_{s(k)=n} |a_k|((M_2 + X)^n - nM_2^{n-1}X - M_2^n),$$

considered as an equation in X.

Proof. If ϵ is small, and $D(\epsilon)$ is chosen as in Lemma 11, the expressions $R(u) - R(0)$ and $A_k(u) - A_k(0)$ appearing in (24) will be small throughout the region $|u| < b - D(\epsilon)$. Also, when b is large, $A_k(0)$ is small whenever $s(k) \neq n$ and near $|a_k|$ whenever $s(k) = n$, (by Lemma 5). From these remarks Lemma 12 follows at once.

Lemma 13. *Equation* (29) *has a non-zero discriminant.*

Proof. Letting Y and r be defined by $Y = X/M_2$, and $r = M_3/M_2^{n-1}\sum_{s(k)=n} |a_k|$, we obtain from (29) the equation

$$(30) \qquad rY = (1 + Y)^n - nY - 1.$$

To prove that the discriminant of (29) is not zero it suffices to prove that (30) and the derived equation

$$(31) \qquad r = n(1 + Y)^{n-1} - n$$

have no common solution. Assuming the contrary, let us suppose that Y is a solution of both (30) and (31). Then

$$(32) \qquad (r + n)Y = (1 + Y)^n - 1,$$

and

$$(33) \qquad (r + n) = n(1 + Y)^{n-1},$$

so that Y is the positive number $r/(r + n)(n - 1)$, and

$$(34) \qquad (1 + Y)^{n-1} = 1 + \gamma, \quad \text{where} \quad \gamma = r/n.$$

Hence $1 + Y = (1 + \gamma)^\tau$, assuming $\tau = (n - 1)^{-1}$. Then, from (32),

$$(n\gamma + n)[(1 + \gamma)^\tau - 1] = (1 + \gamma)^{n\tau} - 1$$

or

$$(35) \qquad n(1 + \gamma)[(1 + \gamma)^\tau - 1] - (1 + \gamma)^{n\tau} + 1 = 0.$$

Now the left-hand member of (35) is zero when $\gamma = 0$, and its derivative with respect to γ is $n(\tau + 1)(1 + \gamma)^\tau - n - n\tau(1 + \gamma)^{n\tau-1}$, hence, (since $n\tau - 1 = \tau$), is the number $n(1 + \gamma)^\tau - n$ which is positive when γ is positive. Thus (35) is false, since γ is positive. This contradiction establishes Lemma 13.

Lemma 14. *There is a positive number* D_0 *such that when* b *is large equation* (24) *has* n *distinct solutions* $Z(u)$, *each analytic in the region* $|u| < b - D_0$. *Thus the series* $\sum_1^\infty Z_j u^j$ *converges in the region* $|u| < b - D_0$ *to an analytic function of* u.

Proof. Since the discriminant of equation (29) is different from zero, and since

the discriminant of a polynomial is a continuous function of the coefficients of the polynomial, it follows from Lemma 12 that when D_0 and b are sufficiently large, the discriminant of (24) will be different from zero throughout the region $|u| < b - D_0$. From this Lemma 14 follows at once.

Part IV. Analytic Solutions of the Approximating q-Difference Equation

Lemma 15. *For all sufficiently large values of b, there are exactly n distinct solutions $y(x) = y^{(t)}(x, b)$, $(t = 1, 2, \cdots, n)$, of equation (12), analytic in the neighborhood of $x = b$. There are positive numbers D and M such that for all sufficiently large values of b, $y^{(t)}(x, b)$, $(t = 1, 2, \cdots, n)$ is analytic in the region $|x - b| < b - D$, and there satisfies the inequality $|y^{(t)}(x, b)| < M|x|^{p/n}$.*

Every function $y^{(t)}(x, b)x^{-p/n}$, $(t = 1, \cdots, n)$, is almost constant.

When b is large, the numbers $y^{(t)}(b, b)b^{-p/n}$ are n distinct numbers, one near each of the numbers $\sigma_1, \sigma_2, \cdots, \sigma_n$, (defined in Lemma 6). (It may be assumed that the superscripts are so assigned to the solutions that $y^{(t)}(b, b)b^{-p/n}$, $(t = 1, \cdots, n)$, tends to σ_t as $b \to \infty$.)

Proof. By Lemma 8 there are exactly n distinct formal power series $\zeta^{(t)} + \sum_1^\infty z_j^{(t)}u^j$, $(t = 1, \cdots, n)$, which satisfy (19). Since $|z_j^{(t)}| \leqq |Z_j|$, $(t = 1, \cdots, n)$, it follows from Lemma 14 that each of these power series converges to a function analytic in the region $|u| < b - D_0$. There are no other solutions of (19), analytic in the neighborhood of $u = 0$.

Let

$$z^{(t)}(x, b) = \zeta^{(t)} + \sum_{j=1}^{\infty} z_j^{(t)}(x - b)^j, \qquad (t = 1, \cdots, n), \tag{36}$$

and let

$$y^{(t)}(x, b) = z^{(t)}(x, b)x^{p/n}, \qquad (t = 1, \cdots, n). \tag{37}$$

Then the functions $z^{(t)}(x, b)$ are n distinct solutions of (15) analytic in $|x - b| < b - D_0$, and the functions $y^{(t)}(x, b)$ are as a consequence n distinct solutions of (12), analytic in $|x - b| < b - D_0$. There are no other solutions of (12), analytic at $x = b$. This establishes the first statement of Lemma 15.

By Lemma 13, equation (29) has n distinct solutions $X_1, X_2, \cdots, X_n$. Let $Z_1(u), Z_2(u), \cdots, Z_n(u)$, be the n distinct solutions of equation (24), analytic in $|u| < b - D_0$, (see Lemma 14).

Because the roots of a polynomial are continuous functions of the coefficients, it follows from Lemma 12 that for every positive ϵ, if $D_2(\epsilon)$ and b are both sufficiently large, with $D_2(\epsilon) > D_0$, there is for every u in the region $|u| < b - D_2(\epsilon)$ and for every w, $(w = 1, 2, \cdots, n)$, a number $X(u, w)$ belonging to the set $X_1, \cdots, X_n$ and satisfying the inequality

$$|Z_w(u) - X(u, w)| < \epsilon. \tag{38}$$

Since $Z_w(u)$ varies continuously with u, and $X(u, w)$ can assume only the discrete values $X_1, X_2, \cdots, X_n$, it follows from (38) that if ϵ is sufficiently small, $X(u, w)$ is a constant function of u. This implies that the subscripts of $Z_1(u), Z_2(u), \cdots, Z_n(u)$ can be chosen in such a way that

$$(39) \qquad | Z_w(u) - X_w | < \epsilon, \qquad (w = 1, 2, \cdots, n),$$

whenever $| u | < b - D_2(\epsilon)$. We assume that this choice of subscripts is made. There is precisely one solution $Z_w(u)$ of (24) which vanishes when $u = 0$. Let us assume that $Z_1(u)$ is this solution. Then from (39) it follows that $| X_1 | < \epsilon$, whence $X_1 = 0$. Hence from (39) again it follows that

$$(40) \qquad | Z_1(u) | < \epsilon, \quad \text{whenever} \quad | u | < b - D_2(\epsilon).$$

Now $Z_1(u)$ is the only solution of (24) which vanishes when $u = 0$, and hence must be identical with the function $\sum_1^\infty Z_j u^j$, since the latter, also, is a solution of (24) which vanishes when $u = 0$. Thus

$$(41) \qquad \left| \sum_1^\infty Z_j u^j \right| < \epsilon \quad \text{when} \quad | u | < b - D_2(\epsilon).$$

Since $| z_j^{(t)} | \leqq Z_j$, $(t = 1, 2, \cdots, n; j = 1, 2, \cdots)$, it follows that

$$(42) \qquad \left| \sum_1^\infty | z_j^{(t)} | \, u^j \right| < \epsilon, \qquad (t = 1, 2, \cdots, n),$$

when $| u | < b - D_2(\epsilon)$.

This implies that $y^{(t)}(x, b)x^{-p/n}$ is almost constant.

By equation (22), $| \zeta^{(t)} | < M_2$. Hence it follows from (42) that $| \zeta^{(t)} + \sum_{j=1}^\infty z_j^{(t)} u^j | < M$, or $| y^{(t)}(x, b) | < M | x |^{p/n}$, when $| x - b | < b - D$, provided $D > D_2(1)$, and $M = M_2 + 1$.

Since $y^{(t)}(b, b)b^{-p/n} = z^{(t)}(b, b) = \zeta^{(t)}$, $(t = 1, 2, \cdots, n)$, the final statement of the lemma follows from Lemma 6.

Part V. Existence of Analytic Solutions of the Difference Equation

Lemma 16. *There exist at* least *n distinct solutions $y^{(t)}(x)$, $(t = 1, 2, \cdots, n)$, of the difference equation* (1), *each analytic in the region $\Re(x) > D$, and there satisfying the inequalities $| y^{(t)}(x) | \leqq M | x |^{p/n}$, $(t = 1, 2, \cdots, n)$. When x tends to ∞ along the positive real axis, $y^{(t)}(x)x^{-p/n}$ tends to σ_t, $(t = 1, 2, \cdots, n)$.*

Proof. Let B be a bounded region which with its boundary is included in the region $\Re(x) > D$. There is a positive number b_0 such that the region $| x - b | < b - D$ will contain B provided $b > b_0$.

Let $\{b_k\}$, $(k = 1, 2, \cdots)$, be a sequence of positive numbers increasing to infinity, with $b_1 > b_0$. Let t be any one of the integers $1, 2, \cdots, n$. The sequence of analytic functions $\{y^{(t)}(x, b_k)\}$, $(k = 1, 2, \cdots)$ is a bounded family in B, because of Lemma 15. Hence there exists a subsequence $\{c_k\}$ of the sequence $\{b_k\}$, such that the sequence of functions $\{y^{(t)}(x, c_k)\}$ converges to a limit func-

tion analytic in B, the convergence being uniform in every closed subset of B.[7] By the standard device of expressing the region $\Re(x) > D$ as the union of a sequence of such regions B, each region including the preceding, the limit function can be continued analytically throughout the region $\Re(x) > D$, so that a function $y^{(t)}(x)$ is obtained, analytic in the region $\Re(x) > D$, and satisfying the relation

(43) $$y^{(t)}(x) = \lim_{k \to \infty} y^{(t)}(x, d_k),$$

where $\{d_k\}$ is some subsequence of $\{b_k\}$, where it is understood that for each value of x in $\Re(x) > D$ the function $y^{(t)}(x, d_k)$ is defined only for sufficiently large values of k, and where the limit is uniform in every closed bounded subset of $\Re(x) > D$.

Now $y^{(t)}(x, b)$ is a solution of (12). Hence, since the q_{ks} appearing in (12) tend to 1 as b tends to infinity, it follows that $y^{(t)}(x)$ is a solution of (1).

Since $| y^{(t)}(x, b) | < M | x |^{p/n}$ in $|x - b| < b - D$, it follows that $| y^{(t)}(x) | \leqq M | x |^{p/n}$ in $\Re(x) > D$.

Let ϵ be any positive number. Since by Lemma 15 $y^{(t)}(x, b)\, x^{-p/n}$ is almost constant, there is a positive number $D(\epsilon)$ such that for all sufficiently large values of b,

(44) $$| y^{(t)}(x, b)x^{-p/n} - y^{(t)}(b, b)b^{-p/n} | < \epsilon$$

when $| x - b | < b - D(\epsilon)$. Hence if x_0 is real and greater than $D(\epsilon)$,

(45) $$| y^{(t)}(x_0 , d_k)x_0^{-p/n} - y^{(t)}(d_k , d_k)d_k^{-p/n} | < \epsilon,$$

for all sufficiently large values of k. By Lemma 15, $\lim y^{(t)}(d_k , d_k)d_k^{-p/n} = \sigma_t$, as $k \to \infty$. By this and (43), it follows from (45) that

(46) $$| y^{(t)}(x_0)x_0^{-p/n} - \sigma_t | \leqq \epsilon.$$

Thus, as x tends to ∞ along the positive real axis, $y^{(t)}(x)x^{-p/n}$ tends to σ_t, $(t = 1, \cdots, n)$. Since the numbers σ_t are distinct, it follows that the functions $y^{(t)}(x)$ are distinct.

Part VI. Approximability, Uniqueness, and Almost Constancy of Solutions of the Difference Equation

Lemma 17. *Let $y_0(x)$ be any solution of the difference equation* (1) *which is analytic in $\Re(x) > D$ and there satisfies the condition $| y_0(x) | < M | x |^{p/n}$, for some positive numbers D and M. Then $y_0(x)$ satisfies the q-difference equation*

(47) $$\sum_{k=1}^{m} a_k(x) \prod_{s=1}^{s(k)} y_0(q_{ks}x + \omega_{ks}) = \psi(x),$$

where

(48) $$\psi(x)x^{-p}$$

[7] Montel, loc. cit.

is almost constant, and $\psi(b)b^{-p} \to c$ as $b \to \infty$, *and the numbers* q_{ks} *are defined as in Definition* 1.

PROOF. By hypothesis

$$\sum_{k=1}^{m} a_k(x) \prod_{s=1}^{s(k)} y_0(x + \omega_{ks}) = \varphi(x).$$

Hence

$$\sum_{k=1}^{m} a_k(x) \prod_{s=1}^{s(k)} y_0(q_{ks} + \omega_{ks}) = \varphi(x) \tag{49}$$
$$+ \sum_{k=1}^{m} a_k(x) \left[\prod_{s=1}^{s(k)} y_0(q_{ks} x + \omega_{ks}) - \prod_{s=1}^{s(k)} y_0(x + \omega_{ks}) \right].$$

Thus we obtain (47), with

$$\psi(x) = \varphi(x) + \sum_{k=1}^{m} a_k(x) \left[\prod_{s=1}^{s(k)} y_0(q_{ks} x + \omega_{ks}) - \prod_{s=1}^{s(k)} y_0(x + \omega_{ks}) \right], \tag{50}$$

and properties (48) to be established. Let

$$d(x) = \psi(x) - \varphi(x). \tag{51}$$

Let

$$e(x) = d(x)x^{-p}. \tag{52}$$

Then

$$e(x) = \sum_{k=1}^{m} a_k(x) x^{-p(n-s(k))/n} h_k(x), \tag{53}$$

where

$$h_k(x) = x^{-ps(k)/n} \left[\prod_{s=1}^{s(k)} y_0(q_{ks} x + \omega_{ks}) - \prod_{s=1}^{s(k)} y_0(x + \omega_{ks}) \right], \quad (k = 1, \cdots, m) \tag{54}$$

By Lemma XI, $h_k(x)$ is almost constant, and $\lim h_k(b) = 0$, as $b \to \infty$.

By (7), if $p \leqq 0$, there are no values of k for which $s(k) \neq n$. Hence, if $p \leqq 0$, $x^{-p(n-s(k))/n} \equiv 1$. If $p > 0$, $x^{-p(n-s(k))/n}$ is bounded as $x \to \infty$, and is almost constant by Lemma III.

By hypothesis, $a_k(x)$ is almost constant, and bounded as $x \to +\infty$.

Hence $\lim_{b\to\infty} e(b) = 0$, and by Lemmas VI, V, $e(x)$ is almost constant.

Now $\psi(x)\, x^{-p} = \varphi(x)\, x^{-p} + e(x)$. By hypothesis $\varphi(x)\, x^{-p}$ is almost constant and $\lim_{b\to\infty} \varphi(b)b^{-p} = c$. Hence $\psi(x)x^{-p}$ is almost constant and $\lim_{b\to\infty} \psi(b)b^{-p} = c$.

LEMMA 18. *Let* $e(x)$ *be the function* $(\psi(x) - \varphi(x))x^{-p}$. *Define* $u = x - b$, $e(x) = \sum_0^\infty e_j\, u^j$, *and*

$$E(u) = \sum_{0}^{\infty} | e_j |\, u^j. \tag{55}$$

Then there is a positive number D_1 such that

$$\lim_{b\to\infty} \max\,(|\,E(u)\,|;\,|\,u\,|\leqq b - D_1) = 0. \tag{56}$$

PROOF. Let $h_k(x)$ be the function defined in (54). Define $h_k(x) = \sum_0^\infty h_{kj}u^j$, $(k = 1, \cdots, m)$. Then by Lemma XI there is a positive number D_1 such that

$$\lim_{b\to\infty} \max\left(\left|\sum_0^\infty |\,h_{kj}\,|\,u^j\right|;\,|\,u\,|\leqq b - D_1\right) = 0. \tag{57}$$

Let $a_k(x) = \sum_0^\infty a_{kj}u^j$. Then, since by hypothesis $a_k(x)$ is almost constant, and $\lim a_k(b) = a_k$, it follows that if D_1 is sufficiently large there is a positive number M such that

$$\left|\sum_0^\infty |\,a_{kj}\,|\,u^j\right| \leqq M, \tag{58}$$

when $|\,u\,| \leqq b - D_1$.

Consider the factor $x^{-p(n-s(k))/n}$ appearing in (53). By (7) the exponent of x is not positive. Hence if we define $\lambda_k(x) = x^{-p(n-s(k))/n}$ with $\lambda_k(x) = \sum_0^\infty \lambda_{kj}u^j$, we have by Lemma III and the fact that $\lambda_k(b)$ is bounded for large b, the result that if D_1 and M are sufficiently large,

$$\left|\sum_0^\infty |\,\lambda_{kj}\,|\,u_j\right| < M, \qquad (k = 1, \cdots, m), \tag{59}$$

when $|\,u\,| < b - D_1$.

Now

$$|\,E(u)\,| \leqq E(|\,u\,|), \tag{60}$$

and by (53)

$$E(|\,u\,|) \leqq \sum_{k=1}^m \left(\sum_{j=0}^\infty |\,a_{kj}\,|\,|\,u\,|^j\right)\left(\sum_{j=0}^\infty |\,\lambda_{kj}\,|\,|\,u\,|^j\right)\left(\sum_{j=0}^\infty |\,h_{kj}\,|\,|\,u\,|^j\right),$$

whence from (57), (58), (59) and (60) follows

$$\lim_{b\to\infty} \max\,(|E(u)\,|;\,|u\,|\leqq b - D_1) = 0.$$

LEMMA 19. *Let $y_0(x)$ be as in Lemma 17. Then $y_0(x)x^{-p/n}$ is almost constant, and there is a value of t such that $\lim_{b\to\infty} y_0(b)b^{-p/n} = \sigma_t$, where $\sigma_1, \cdots, \sigma_n$ are the numbers defined in Lemma 6.*

PROOF. By Lemma 17, $y_0(x)$ satisfies equation (47). Since the hypotheses for the coefficients of (47) are identical with those for the coefficients of equation (12), and since the numbers $\sigma_1, \cdots, \sigma_n$ are defined in terms of the numbers c and a_k, $(k = 1, \cdots, m)$ which have the same significance for (47) as for (12), it follows that Lemma 15 remains valid with reference to the solution $y_0(x)$ of equation (47), instead of to the solution $y^{(t)}(x, b)$ of equation (12).

Hence $y_0(x)x^{-p/n}$ is almost constant, and if b is large, $y_0(b)b^{-p/n}$ is near one of the numbers $\sigma_1, \cdots, \sigma_n$. Which one of these $y_0(b)b^{-p/n}$ is near might conceivably depend upon b, but $y_0(b)b^{-p/n}$ varies continuously with b, and therefore there is a fixed value of t such that $y_0(b)b^{-p/n}$ is near σ_t when b is large.

LEMMA 20. *Let $y_0(x)$ be as in Lemma 17. Let t be the number exhibited in Lemma 19, such that* $\lim y_0(b)b^{-p/n} = \sigma_t$ *as* $b \to \infty$. *Let $y^{(t)}(x, b)$ be the function described in Lemma 15. Then there is a positive number D_2 such that*

$$y_0(x) = \lim_{b\to\infty} y^{(t)}(x, b), \tag{61}$$

when $\Re(x) > D_2$, the limit being uniform is every closed bounded subset of the region $\Re(x) > D_2$.

PROOF. Define $z_0(x) = y_0(x)x^{-p/n}$, and $z(x) = y^{(t)}(x, b)x^{-p/n}$. Since $y^{(t)}(x, b)$ satisfies (12), $z(x)$ satisfies (15). Likewise, since $y_0(x)$ satisfies (47), $z_0(x)$ satisfies

$$\sum_{k=1}^{m} \alpha_k(x) \prod_{s=1}^{s(k)} z_0(q_{ks}x + \omega_{ks}) = \rho(x) \tag{62}$$

where $\rho(x) = \psi(x)x^{-p}$, and the $\alpha_k(x)$ are defined by (16).

Subtraction of equation (15) from equation (62) gives

$$\sum_{k=1}^{m} \alpha_k(x) \left[\prod_{s=1}^{s(k)} z_0(q_{ks}x + \omega_{ks}) - \prod_{s=1}^{s(k)} z(q_{ks}x + \omega_{ks})\right] = e(x) \tag{63}$$

where $e(x)$ is defined by (51) and (52).

Hence

$$\sum_{k=1}^{m} \sum_{s=1}^{s(k)} c_{ks}(x)[z_0(q_{ks}x + \omega_{ks}) - z(q_{ks}x + \omega_{ks})] = e(x), \tag{64}$$

where

$$c_{ks}(x) = \alpha_k(x) \prod_{i=1}^{s-1} z(q_{ki}x + \omega_{ki}) \prod_{i=s+1}^{s(k)} z_0(q_{ki}x + \omega_{ki}). \tag{65}$$

Let (66) $v(x) = z_0(x) - z(x)$. Then

$$\sum_{k=1}^{m} \sum_{s=1}^{s(k)} c_{ks}(x)v(q_{ks}x + \omega_{ks}) = e(x). \tag{67}$$

Let $c_{ks}(x) = \sum_{j=0}^{\infty} c_{ksj}u^j$, $v(x) = \sum_{j=0}^{\infty} v_j u^j$, $e(x) = \sum_{j=0}^{\infty} e_j u^j$, where $u = x - b$. Then, by (67) and Lemma 1,

$$\sum_{k=1}^{m} \sum_{s=1}^{s(k)} \left(\sum_{j=0}^{\infty} c_{ksj}u^j\right)\left(\sum_{j=0}^{\infty} v_j q_{ks}^j u^j\right) = \sum_{j=0}^{\infty} e_j u^j, \tag{68}$$

and therefore

$$\sum_{k=1}^{m}\sum_{s=1}^{s(k)} c_{ks0}\sum_{j=0}^{\infty} v_j q_{ks}^j u^j = \sum_{j=0}^{\infty} e_j u^j - \sum_{k=1}^{m}\sum_{s=1}^{s(k)}\left(\sum_{j=1}^{\infty} c_{ksj} u^j\right)\left(\sum_{j=0}^{\infty} v_j q_{ks}^j u^j\right). \tag{69}$$

The coefficient of u^j in the left-hand member of (69) is $v_j d_j$, where, by definition,

$$d_j = \sum_{k=1}^{m}\sum_{s=1}^{s(k)} c_{ks0}\, q_{ks}^j\,. \tag{70}$$

Now from (65) follows

$$c_{ks0} = c_{ks}(b) = \alpha_k(b)\prod_{i=1}^{s-1} z(b)\prod_{i=s+1}^{s(k)} z_0(b) = \alpha_k(b)\prod_{i=1}^{s-1} y^{(t)}(b, b)b^{-p/n}\prod_{i=s+1}^{s(k)} y_0(b)b^{-p/n}.$$

Hence, by Lemma 5, if $s(k) \neq n$, $c_{ks0} \to 0$ as $b \to \infty$. Therefore, when b is large, all the terms in d_j for which $s(k) \neq n$ are small. The remaining terms are near $(\sigma_t)^{n-1}\sum_{s(k)=n}\sum_{s=1}^{n} a_k q_{ks}^j$, the modulus of which is at least $M_1^{n-1}\,\Re(\sum_{s(k)=n}\sum_{s=1}^{n} a_k q_{ks}^j)$, and this, by Lemma 5, is bounded below by a positive number independent of j and b. Thus

$$|\, d_j \,| > M_4\,, \tag{71}$$

where M_4 is a positive number independent of j. Let $C_{ksj} = |\, c_{ksj}\,|$, $E_j = |\, e_j\,|$, $(k = 1, \cdots, m;\ s = 1, \cdots, s(k);\ j = 0, 1, 2, \cdots)$. Let V_j $(j = 0, 1, 2, \cdots)$ be defined by

$$\sum_{j=0}^{\infty} M_4 V_j u^j = \sum_{j=0}^{\infty} E_j u^j + \sum_{k=1}^{m}\sum_{s=1}^{s(k)}\left(\sum_{j=1}^{\infty} C_{ksj} u^j\right)\left(\sum_{j=0}^{\infty} V_j u^j\right). \tag{72}$$

Then

$$|\, v_j \,| \leqq V_j\,,\ (j = 0, 1, 2, \cdots), \tag{73}$$

by (69) and (71). (A formal proof may be given by induction on j.)

Define $V(u) = \sum_0^\infty V_j u^j$, $E(u) = \sum_0^\infty E_j u^j$, $C_{ks}(u) = \sum_{j=0}^{\infty} C_{ksj} u^j$, $(k = 1, \cdots, m;\ s = 1, \cdots, s(k)$. Then from (72) it follows that

$$M_4 V(u) = E(u) + \left[\sum_{k=1}^{m}\sum_{s=1}^{s(k)} (C_{ks}(u) - C_{ks}(0))\right] V(u), \tag{74}$$

or

$$V(u) = E(u) \Big/ \left[M_4 - \sum_{k=1}^{m}\sum_{s=1}^{s(k)} (C_{ks}(u) - C_{ks}(0))\right]. \tag{75}$$

Now the function $z(x)$ is almost constant, by Lemma 15. Thus, if $z(x) = \sum_0^\infty z_j u^j$, there is for every positive ϵ a positive $D(\epsilon)$ such that for all sufficiently large values of b,

$$\left|\sum_0^\infty |\, z_j \,|\, u^j - |\, z_0 \,|\right| < \epsilon, \tag{76}$$

when $|u| < b - D(\epsilon)$. Now $z(q_{ks}x + w_{ks}) = \sum_0^\infty z_j q_{ks}^j u^j$, $(k = 1, \cdots, m;$ $s = 1, \cdots, s(k))$, by Lemma 1, and it follows from (76) that $|\sum_0^\infty |z_j q_{ks}^j| u^j - |z_0|| < \epsilon$, when $|u| < b - D(\epsilon)$. Hence $z(q_{ks}x + \omega_{ks})$ is almost constant, $(k = 1, \cdots, m; s = 1, \cdots, s(k))$.

Likewise the function $z_0(x)$ is almost constant, since $y_0(x)$ is a solution of (47), and therefore every function $z_0(q_{ks}x + \omega_{ks})$ is almost constant.

By (28), $\alpha_k(x)$ is almost constant $(k = 1, \cdots, m)$.

Hence, by Lemma VI, $c_{ks}(x)$ is almost constant, $(k = 1, \cdots, m; s = 1, \cdots, s(k))$.

This implies that there is a positive number D_2 such that

$$\left| \sum_{k=1}^{m} \sum_{s=1}^{s(k)} (C_{ks}(u) - C_{ks}(0)) \right| < \frac{1}{2} M_4, \text{ if } |u| \leqq b - D_2 .$$

Thus (75) implies

$$|V(u)| < 2 |E(u)| M_4^{-1} \text{ when } |u| \leqq b - D_2. \tag{77}$$

Hence from Lemma 18 follows max $(|V(u)|; |u| \leqq b - D_2) \to 0$ as $b \to \infty$, provided D_2 is taken greater than D_1.

Therefore from (73) follows max $(|v(x)|; |x - b| \leqq b - D_2) \to 0$ as $b \to \infty$ or max $(|z_0(x) - z(x)|; |x - b| \leqq b - D_2) \to 0$ as $b \to \infty$.

This implies that (61) is valid with the limit uniform in every closed bounded subset of the region $\Re(x) > D_2$.

LEMMA 21. *If D is any positive number there are at* most *n distinct solutions of the difference equation* (1), *analytic in the region $\Re(x) > D$, and there satisfying* (9).

PROOF. Assume that there are more than n such solutions, say $y_1(x), y_2(x), \cdots, y_l(x)$, with $l > n$. Then by Lemma 20 there is a number $t(i)$ in the set $(1, 2, \cdots, n)$ such that

$$y_i(x) = \lim y^{(t(i))}(x, b) \text{ as } b \to \infty. \tag{78}$$

But there must be two values of i for which $t(i)$ has the same value, and this implies by (78) that two of the functions $y_i(x)$ are identical. This contradiction establishes Lemma 21.

PART VII. SUMMARY

Theorem 2 follows immediately from Lemmas 16, 19, 20, and 21.

Appendix. Almost Constant Functions

1. DEFINITION OF ALMOST CONSTANT FUNCTIONS. See Definition 2, above.

2. NOTATION. *If f is analytic at $x = b$, and can therefore be expanded in a Taylor's series $\sum_{j=0}^\infty \beta_j(x - b)^j$, the function $\sum_{j=0}^\infty |\beta_j| u^j$ will be denoted by the symbol $f_A(u, b)$.*

3. LEMMA I. *Let $f(x)$ be analytic in a half-plane $\Re(x) > D > 0$, and satisfy there the inequality $|f(x) - f(b)| \leqq M(b^{-1} + |x|^{-1})$ for some M and all sufficiently large values of b. Then $f(x)$ is almost constant.*

PROOF. Let $f(x) = \sum_{j=0}^{\infty} \beta_j(x - b)^j$. Let D_0 be any number greater than D. Let b be any number greater than D_0. Let $|u| \leqq b - 2D_0$, $\rho = b - D_0$. Then

$$
\begin{aligned}
|f_A(u, b) - f_A(0, b)|^2 &= \left|\sum_{j=1}^{\infty} |\beta_j| u^j\right|^2 = \left|\sum_{j=1}^{\infty} |\beta_j| \rho^j(u/\rho)^j\right|^2 \\
&\leqq \left(\sum_{j=1}^{\infty} |\beta_j|^2 \rho^{2j}\right)\left(\sum_{j=1}^{\infty} |u/\rho|^{2j}\right) \\
&= \left[(2\pi)^{-1} \int_0^{2\pi} |f(b + \rho e^{i\vartheta}) - f(b)|^2 d\theta\right][|u|^2/(\rho^2 - |u|^2)]^{(8)} \\
&\leqq \left[(2\pi)^{-1} \int_0^{2\pi} M^2(b^{-2} + 2b^{-1}|b + \rho e^{i\vartheta}|^{-1} + |b + \rho e^{i\vartheta}|^{-2})\, d\theta\right] bD_0^{-1} \\
&\leqq \left[2(2\pi)^{-1} \int_0^{2\pi} M^2(b^{-2} + |b + \rho e^{i\vartheta}|^{-2})\, d\theta\right] bD_0^{-1} \\
&= 2M^2[b^{-2} + (b^2 - \rho^2)^{-1}]\, bD_0^{-1} \leqq 2M^2[b^{-1} + D_0^{-1}]\, D_0^{-1}.
\end{aligned}
$$

If D_0 is large, the last member is small, for all large values of b. Hence $f(x)$ is almost constant.

4. LEMMA II. *Every function of the form* $(\alpha + \beta/x)^s$, *where* α *is positive,* β *is non-negative,* s *is real, and the branch chosen (in the right half-plane) is positive for positive* x, *is almost constant.*

PROOF. Let $f(x) = (\alpha + \beta/x)^s$. Let x be any number such that $|\beta/x| < \alpha/2$. Then $f(x) - \alpha^s = (\alpha + \beta/x)^s - \alpha^s = (\alpha + \rho e^{i\vartheta})^s - \alpha^s$, where a non-negative ρ and a real ϑ are defined by $\rho e^{i\vartheta} = \beta/x$. Let $F(\rho) = (\alpha + \rho e^{i\vartheta})^s$. Then $f(x) - \alpha^s = F(\rho) - F(0) = \int_0^\rho F'(t)dt = \int_0^\rho s(\alpha + te^{i\vartheta})^{s-1} e^{i\vartheta}dt.$

Therefore $|f(x) - \alpha^s| \leqq \rho M \leqq M_1 |x|^{-1}$ for suitable constants M, M_1 independent of ρ, $|x|$ respectively. Likewise $|f(b) - \alpha^s| \leqq M_1 b^{-1}$. Hence $|f(x) - f(b)| \leqq M_1(b^{-1} + |x|^{-1})$. By Lemma I, $f(x)$ is almost constant.

5. LEMMA III. *Every function of the form* $(x - a)^{-\sigma}$ *where* a *is an arbitrary complex number, where* σ *is non-negative, and where the branch chosen (in the half-plane* $\Re(x) > \Re(a)$*) is any branch, is almost constant.*

PROOF. We may and do assume that $\sigma > 0$. Let $f(x) = (x - a)^{-\sigma}$. Let $b > |a|$. Let $c = b - a$. Then $f(x) = (x - b + c)^{-\sigma} = c^{-\sigma}(1 + (x - b)/c)^{-\sigma} = c^{-\sigma}\sum_{j=0}^{\infty} [(-\sigma)(-\sigma - 1) \cdots (-\sigma - j + 1)] (x - b)^j c^{-j}/j!$
Hence $f_A(u, b) = |c|^{-\sigma} \sum_{j=0}^{\infty} |(-\sigma)(-\sigma - 1) \cdots (-\sigma - j + 1)| u^j |c|^{-j}/j! = |c|^{-\sigma} \sum_{j=0}^{\infty} (-1)^j(-\sigma)(-\sigma - 1) \cdots (-\sigma - j + 1)u^j |c|^{-j}/j! = (|c| - u)^{-\sigma}$. Hence $f_A(u, b) - f_A(0, b) = (|c| - u)^{-\sigma} - |c|^{-\sigma}$. Hence if $|u| < b - D$, where D is any number greater than $|a|$, $|f_A(u, b) - f_A(0, b)| < (D - |a|)^{-\sigma} + (b - |a|)^{-\sigma}$. If ϵ is any positive number, D can be chosen so that

[8] Landau, "Darstellung und Begrundung einiger neuerer Ergebnisse der Funktionentheorie," Berlin 1929, page 8.

$(D - |a|)^{-\sigma} < \epsilon/2$, and then for all sufficiently large values of b, $|f_A(u, b) - f_A(0, b)| < \epsilon$. Hence $f(x)$ is almost constant.

6. Lemma IV. *If $f(x)$ is almost constant, $f(b)$ tends to a limit as b tends to positive infinity.*

Proof. Let ϵ be any positive number. Let $D(\epsilon)$ be such that $|f_A(u, b) - f_A(0, b)| < \epsilon$ when $|u| < b - D(\epsilon)$, provided $b > b_1$. Choose b_2 greater than b_1 and greater than $D(\epsilon)$, and let $b_3 \geqq b_2$. Then $|f_A(u, b_3) - f_A(0, b_3)| < \epsilon$, if $|u| < b_3 - D(\epsilon)$. *A fortiori*, $|f(x) - f(b_3)| < \epsilon$ if $|x - b_3| < b_3 - D(\epsilon)$. In particular, $|f(b_2) - f(b_3)| < \epsilon$, since $|b_2 - b_3| < b_3 - D(\epsilon)$. Hence $\lim f(b)$, as $b \to \infty$, exists.

7. Lemma V. *If $f_1(x, b), \cdots, f_N(x, b)$ are almost constant, so is $f_1(x, b) + \cdots + f_N(x, b)$.* Proof omitted.

8. Lemma VI. *If $f_1(x, b), \cdots, f_N(x, b)$ are almost constant, so is the product $f_1(x, b)f_2(x, b) \cdots f_N(x, b)$,* provided $|f_1(b, b), \cdots, |f_N(b, b)|$ are bounded by a number M.

Proof. Case 1. $N = 2$. Let $f(x, b)$, $g(x, b)$ be almost constant. Let $h(x, b) = f(x, b)g(x, b)$. To prove $h(x, b)$ almost constant.

Let ϵ be positive, and let $D(\epsilon)$ be such that $|f_A(u, b) - f_A(0, b)| < \epsilon$ and $|g_A(u, b) - g_A(0, b)| < \epsilon$, when $|u| < b - D(\epsilon)$, (for all sufficiently large values of b). Then $|h_A(u, b) - h_A(0, b)| \leqq h_A(|u|, b) - h_A(0, b) \leqq f_A(|u|, b)g_A(|u|, b) - f_A(0, b)g_A(0, b) = [f_A(|u|, b) - f_A(0, b)][g_A(|u|, b) - g_A(0, b)] + f_A(0, b)[g_A(|u|, b) - g_A(0, b)] + g_A(0, b)[f_A(|u|, b) - f_A(0, b)] < \epsilon^2 + M\epsilon + M\epsilon$. This quantity can be made small by taking ϵ small. Hence $h(x, b)$ is almost constant.

Case 2. $N \geqq 2$. Proof by induction.

9. Lemma VII. *Every rational function whose numerator-degree is not greater than its denominator-degree, is almost constant.*

Proof. Let $f(x)$ be such a function. Then, since $f(x)$ can be expanded in partial fractions, there is a relation $f(x) = c_0 + \sum_{j,k} c_{jk}(x - a_k)^{-j}$, where the a_k are the poles of $f(x)$ and j is always positive. Now the constants c_0, c_{jk} are obviously almost constant, and each of the functions $(x - a_k)^{-j}$ is almost constant, by Lemma III, so by Lemmas VI and V, $f(x)$ is almost constant.

10. Lemma VIII. *Let $f(x)$ be analytic in $\Re(x) > D > 0$, and there satisfy the inequality $|f(x)| < M|x|^p$ for some positive M and some real p. Then there is a positive constant M_1 such that $|f'(x)| < M_1|x|^p[\Re(x) - D]^{-1}$ in $\Re(x) > D$.*

Proof. Let x be any complex number such that $\Re(x) > D$. Let C be the circle $\zeta = x + \rho e^{i\vartheta}$, where $\rho = (\Re(x) - D)/2$. Then $f'(x) = (2\pi i)^{-1}\int_C f(\zeta)(\zeta - x)^{-2}d\zeta$. Hence $|f'(x)| \leqq (2\pi)^{-1} M \cdot M(x)\rho^{-2}\, 2\pi\rho = M\, M(x)\rho^{-1}$, where $M(x)$ is the maximum over C of the function $|\zeta|^p$. Now on C, $|\zeta| > |x| - |x|/2$, and $|\zeta| < |x| + |x|/2$. Hence, if $p \geqq 0$, $|\zeta|^p \leqq (3/2)^p|x|^p$, and if $p \leqq 0$, $|\zeta|^p \leqq (1/2)^p|x|^p$. Thus there is an M_1 such that $|f'(x)| < M_1|x|^p[\Re(x) - D]^{-1}$.

11. Lemma IX. *Let ω, D be positive numbers such that $D > \omega$. Let $f(x)$ be*

analytic in $\Re(x) > D$, *and there satisfy the inequality* $|f'(x)| < M_1 |x|^p (\Re(x) - D)^{-1}$ *for some positive* M_1. *Then there is a positive number* M_2 *such that if* b *is any number greater than* 3ω, *and if* $q = 1 - \omega/b$, *then the inequality* $|f(x + \omega) - f(qx + \omega)| < M_2 (1 - q) |x|^{p+1}/\Re(x)$ *is valid for all* x *such that* $\Re(x) > 2D$.

PROOF. $\Re(qx + \omega) = q\Re(x) + \omega > (1 - 1/3)(2D) + \omega > D$. Hence $qx + \omega$ belongs to the half-plane $\Re(x) > D$. Then $f(x + \omega) - f(qx + \omega) = \int_{qx+\omega}^{x+\omega} f'(\zeta)d\zeta$, where the integral is taken along the straight line segment joining $qx + \omega$ to $x + \omega$. Thus $|f(x + \omega) - f(qx + \omega)| \leqq |(x + \omega) - (qx + \omega)| M'$, where M' is the maximum value of $|f'(\zeta)|$ on the line segment. Thus for some number λ between q and 1

$|f(x + \omega) - f(qx + \omega)| \leqq (1 - q) |x| M_1 |\lambda x + \omega|^p (\Re(\lambda x + \omega) - D)^{-1} < M_2(1 - q) |x|^{p+1}[\Re(x)]^{-1}$ for some suitable M_2.

12. LEMMA X. *Let* M, D *be positive numbers. Let* $h(x, b)$ *be a function of* x *and* b *which, for all sufficiently large values of* b, *is analytic in* $\Re(x) > D$, *and there satisfies the inequality* $|h(x, b)| \leqq Mb^{-1} |x|/\Re(x)$. *Let the Taylor's series for* $h(x, b)$ *in powers of* $x - b$ *be* $\sum_{n=0}^{\infty} h_n(b) (x - b)$." *Then* $|\sum_{n=0}^{\infty} |h_n(b)| u^n| \leqq M D^{3/4}b^{1/4}$ *when* $|u| \leqq b - 2D$, (*for all sufficiently large values of* b).

PROOF. Let b be any number greater than $3D$. Let ρ equal $b - D$, and let u satisfy the inequality $|u| \leqq b - 2D$. Then

$$\left|\sum_{n=0}^{\infty} |h_n(b)| u^n\right|^2 = \left|\sum_{n=0}^{\infty} |h_n(b)|\rho^n(u/\rho)^n\right|^2 \leqq \left(\sum_{n=0}^{\infty} |h_n(b)|^2 \rho^{2n}\right)\cdot\left(\sum_{n=0}^{\infty} |u/\rho|^{2n}\right)$$

$$= \left[(2\pi)^{-1} \int_0^{2\pi} |h(b + \rho e^{i\theta}, b)|^2 d\theta\right](1 - |u|^2/\rho^2)^{-1}$$

$$\leqq \left[(2\pi)^{-1} \int_0^{2\pi} M^2 b^{-2} |b + \rho e^{i\theta}|^2 [\Re(b + \rho e^{i\theta})]^{-2} d\theta\right]$$

$$\times \rho^2(\rho^2 - |u|^2)^{-1}.$$

Now $b^{-2}\rho^2 \leqq 1$, and $\rho^2 - |u|^2 \geqq (b - D)^2 - (b - 2D)^2 = 2bD - 3D^2 > bD$. Hence

$$\left|\sum_{n=0}^{\infty} |h_n(b)| u^n\right|^2 \leqq (2\pi)^{-1}M^2\left[\int_0^{2\pi} |b + \rho e^{i\theta}|^2 [\Re(b + \rho e^{i\theta})]^{-2} d\theta\right]b^{-1} D^{-1}$$

$$= (2\pi bD)^{-1} M^2 2\pi b(b^2 - \rho^2)^{-\frac{1}{2}} = M^2 D^{-1}(2bD - D^2)^{-\frac{1}{2}}.$$

Hence

$$\left|\sum_{n=0}^{\infty} |h_n(b)| u^n\right| \leqq MD^{-\frac{3}{4}}b^{-\frac{1}{4}}.$$

13. LEMMA XI. *Let* $f(x)$ *be analytic in* $\Re(x) > D > 0$, *and there satisfy the inequality* $|f(x)| \leqq M |x|^\sigma$ *for some positive* M *and some real* σ. *Let* $\omega_1, \omega_2, \cdots, \omega_s$ *be non-negative numbers, and let* $q_1, q_2, \cdots, q_s$ *be defined by* $q_j = 1 - \omega_j b^{-1}$, $(j = 1, \cdots, s)$. *Then the function* $h(x)$, (*of* x *and* b), *defined by* $h(x) = x^{-\sigma s}$

($\prod_{j=1}^{s} f(x + \omega_j) - \prod_{j=1}^{s} f(q_j x + \omega_j)$) *is almost constant and* $h(b) \to 0$, as $b \to \infty$. *Moreover, if* $h(x) = \sum_{n=0}^{\infty} h_n(x - b)^n$, *and if* $H(x) = \sum_{n=0}^{\infty} | h_n | (x - b)^n$, *then there is a positive number* D' *such that max* $(| H(x) |; | x - b | \leqq b - D') \to 0$ as $b \to \infty$.

Proof. By Lemmas VIII, IX, there is a positive constant M_2 such that $| f(x + \omega_t) - f(q_t x + \omega_t) | < M_2 \omega_t b^{-1} | x |^{\sigma+1}/\Re(x)$ for all sufficiently large values of b, provided $\Re(x) > 2D$, ($t = 1, 2, \cdots, s$). Now

$$h(x) = x^{-\sigma s} \sum_{t=1}^{s} [f(x + \omega_t) - f(q_t x + \omega_t)] \prod_{j=1}^{t-1} f(x + \omega_j) \prod_{j=t+1}^{s} f(q_j x + \omega_j).$$

Hence by Lemma IX,

$$| h(x) | \leqq | x |^{-\sigma s} \sum_{t=1}^{s} M_2 \omega_t b^{-1} | x |^{\sigma+1} [\Re(x)]^{-1} \prod_{j=1}^{t-1} M | x + \omega_j |^{\sigma} \prod_{j=t+1}^{s} M | q_j x + \omega_j |^{\sigma}$$

$\leqq M_3 b^{-1} | x |/\Re(x)$, when $\Re(x) > D'$, for some suitably chosen positive numbers M_3 and D'. Hence Lemma X is applicable, to prove that $h(x)$ is almost constant, that $h(b) \to 0$ as $b \to \infty$, and that max $(| H(x) |; | x - b | \leqq b - D') \to 0$, as $b \to \infty$.

Columbia University.

ANNALS OF MATHEMATICS
Vol. 44, No. 3, July, 1943

A NOTE ON CESARO SUMMABILITY OF FOURIER SERIES

BY FU TRAING WANG

(Received December 28, 1942)

1. It is known that the (C, α) continuity of a function is not sufficient for the (C, α) summability of its Fourier series at a point. Many sufficient conditions for (C, α) summability of a Fourier series were found by many authors.[1] The object of this note is also to find a sufficient condition which arises from a convergence criterion for Fourier series recently proved by the author.[2]

Let $\phi(t)$ be an even integrable periodic function with period 2π and its Fourier series be

$$\phi(t) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nt, \tag{1.1}$$

and let us suppose

$$\phi_\beta(t) = \frac{1}{\Gamma(\beta)} \int_0^t (t - u)^{\beta-1}\{\phi(u) - s\}\, du = o(t^\gamma), \qquad (\gamma > \beta > 0). \tag{1.2}$$

Then we have the following convergence criterion.

If (1.2) *holds and* $a_n > - K_n^{-\frac{\beta}{\gamma}}$ *then the Fourier series* (1.1) *converges to sum* s *at* $t = 0$.

Without the order condition in this convergence criterion we have the following

THEOREM. *Let n be a positive integer such that $n \geqq \beta > n - 1$. If* (1.2) *holds then the Fourier series* (1.1) *is summable* $(C, (\gamma(n - 1) + \beta)/(\gamma + n - \beta))$ *to sum s at $t = 0$.*

Put $\alpha = (\gamma(n - 1) + \beta)/(\gamma + n - \beta)$; then we can easily verify the following inequality:

$$\beta > \alpha > n - 1.$$

It is evident that the theorem is not true for the case $\beta = \gamma$.

2. In order to prove the theorem we put

$$C_\alpha(\omega) = \frac{a_0}{2}\omega^\alpha + \sum_{n<\omega} (\omega - n)^\alpha a_n$$

and

$$\gamma_\alpha(\omega, t) = \int_0^\omega (\omega - x)^\alpha \cos xt\, dx$$

[1] Hardy and Littlewood [3]. Gergen [4].

[2] Wang [5].

Then we get[3]

$$(2.1)\qquad C_\alpha(\omega) = \frac{2}{\pi}\int_0^\pi \phi(t)\gamma_\alpha(\omega, t)\,dt + s\omega^\alpha + o(\omega^\alpha).$$

Concerning the function $\gamma_\alpha(\omega, t)$ we have the following properties.

LEMMA 1.[4] *Let K_i be certain constants depending only on β; then*

$$(2.2.1)\qquad \gamma_\alpha^{(n)}(\omega, t) = \frac{\partial^n}{\partial t^n}\{\gamma_\alpha(\omega, t)\} = \alpha\cdots(\alpha - n + 1)\omega^n t^{-n}\gamma_{\alpha-n}(\omega, t) + \sum_{r=0}^{n-1} k_r\,\omega^r t^{-n}\gamma_{\alpha-r}(\omega, t)$$

and

$$(2.2.2)\qquad \gamma_{\alpha-n}(\omega,t) = O(\omega^{\alpha-n+1}) \quad \textit{for all } \omega \textit{ and } t$$

and

$$(2.2.3)\qquad \gamma_{\alpha-n}(\omega,t) = \Gamma(\alpha - n + 1)\,\frac{\cos\left[\omega t - \frac{\pi}{2}(\alpha - n + 1)\right]}{t^{\alpha-n+1}} + O(\omega^{\alpha-n-1}t^{-2})$$

for $\omega t \geqq 1$.

LEMMA 2. *Let* $\omega' = \frac{\omega}{\rho}$ *and* $\eta = \frac{\alpha - n + 1}{\alpha + 1}$; *if* $\phi_n(t) = o(t^n)$ *then*

$$(2.3)\qquad C_\alpha(\omega) = (-1)^n\frac{2}{\pi}\omega^n\int_0^{\omega'^{-\eta}} \phi_n(t)\,\frac{\cos\left[\omega t - \frac{\pi}{2}(\alpha - n + 1)\right]}{t^{\alpha+1}}\,dt + O(\rho^{-\eta(\alpha+1)}\omega^\alpha) + s\omega^\alpha + o(\omega^\alpha).$$

PROOF. From integration by parts and (2.1) we get

$$(2.4.1)\qquad C_\alpha(\omega) = (-1)^n \pm \frac{2}{\pi}\int_0^\pi \phi_n(t)\gamma_\alpha^{(n)}(\omega, t)\,dt + s\omega^\alpha + o(\omega^\alpha),$$

since

$$\int_0^\pi \phi_n(t)\,\frac{\gamma_{\alpha-i}(\omega,t)}{t^n}\,dt = o(\omega^{\alpha-i}) \quad \text{for} \quad 1 \leqq i \leqq n.$$

By Lemma 1 and (2.4.1) we get

$$(2.4.2)\qquad C_\alpha(\omega) = (-1)^n\frac{2}{\pi}\Gamma(\alpha)\omega^n\int_{\omega^{-1}}^\pi \phi_n(t)\,\frac{\cos\left[\omega t - \frac{\pi}{2}(\alpha - n + 1)\right]}{t^{\alpha+1}}\,dt + s\omega^\alpha + o(\omega^\alpha).$$

---

[3] Gergen [4].

[4] Bosanquet and Linfoot [2], and Bosanquet [1].

By the second mean value theorem then

$$\int_{\omega'^{-\eta}}^{\pi} \phi_n(t) \frac{\cos\left[\omega t - \frac{\pi}{2}(\alpha - n + 1)\right]}{t^{\alpha+1}} dt = O(\rho^{-\eta(\alpha+1)}\omega^{\alpha-n}) \tag{2.4.3}$$

From (2.4.2) and (2.4.3), Lemma 2 follows.

LEMMA 3. *Put*

$$E(\omega, u) = \int_u^{\omega'^{-\eta}} (t-u)^{\eta-\beta-1} \frac{\cos\left[\omega t - \frac{\pi}{2}(\alpha - n + 1)\right]}{t^{\alpha+1}} dt;$$

then

$$E(\omega, u) = O(u^{n-\beta-\alpha-1}) \qquad \textit{for all } \omega \textit{ and } u, \tag{2.5.1}$$

and

$$E(\omega, u) = \Gamma(n-\beta)\omega^{\beta-n} \frac{\cos\left[\omega u - \frac{\pi}{2}(\alpha + \beta - 2n + 1)\right]}{u^{\alpha+1}} + O\{(\omega'^{-\eta} - u)^{n-\beta-1}\omega^{\alpha-n})\} + O(u^{n-\alpha-\beta-1}\omega^{-1}) \tag{2.5.2}$$

for $wu \geqq 1$.

PROOF. By the second mean value theorem we get

$$\int_{\omega'^{-\eta}}^{\infty} (t-u)^{n-\beta-1} \frac{\cos\left[\omega t - \frac{\pi}{2}(\alpha - n + 1)\right]}{t^{\alpha+1}} dt = O\{(\omega'^{-\eta} - u)^{n-\beta-1}\omega^{\alpha-n}\}. \tag{2.6.1}$$

From changes of variable and the second mean value theorem we then obtain

$$\begin{aligned} &\int_u^{\infty} (t-u)^{n-\beta-1} \frac{\cos\left[\omega t - \frac{\pi}{2}(\alpha - n + 1)\right]}{t^{\alpha+1}} dt \\ &= u^{n-\alpha-\beta-1}\int_0^{\infty} \nu^{n-\beta-1} \cos\left[\omega u(1+\nu) - \frac{\pi}{2}(\alpha - n + 1)\right] d\nu \\ &\qquad + O(u^{n-\alpha-\beta-2}\omega^{-1}). \end{aligned} \tag{2.6.2}$$

By a theorem on Γ-functions[5] we get

$$\begin{aligned} &\int_0^{\infty} \nu^{n-\beta-1} \cos\left[\omega u(1+\nu) - \frac{\pi}{2}(\alpha - n + 1)\right] d\nu \\ &= \Gamma(n-\beta) \frac{\cos\left[\omega u - \frac{\pi}{2}(\alpha + \beta - 2n + 1)\right]}{(u\omega)^{n-\beta}}. \end{aligned} \tag{2.6.3}$$

[5] Titchmarsh [6].

A combination of (2.6.1), (2.6.2), (2.6.3) will give the proof of the lemma.

PROOF OF THE THEOREM. If (1.2) holds then

$$\phi_\beta(t) = o(t^\beta). \quad \text{Hence} \quad \phi_n(t) = o(t^n).$$

and

$$\phi_n(t) = \frac{1}{\Gamma(n-\beta)} \int_0^t (t-u)^{n-\beta-1} \phi_\beta(u)\, du. \tag{2.7}$$

If we substitute (2.7) in (2.3) and invert the order of repeated integration we have, by Lemma 3,

$$C_\alpha(\omega) = (-1)^n \frac{2}{\pi} \Gamma(\alpha)\omega^\beta \int_0^{\omega'^{-\eta}} \phi_\beta(u) E(\omega, u)\, du + O(\rho^{-\eta(\alpha+1)}\omega^\alpha) + s\omega^\alpha + o(\omega^\alpha).$$

By (2.5.1) and (2.5.2) we have

$$C_\alpha(\omega) = (-1)^n \frac{2}{\pi} \Gamma(\alpha)\omega^\beta \int_{\omega^{-1}}^{\omega'^{-\eta}} \phi_\beta(u) \frac{\cos\left[\omega u - \frac{\pi}{2}(\alpha+\beta-2n+1)\right]}{u^{\alpha+1}}\, du + O(\rho^{-\eta(\alpha+1)}\omega^\alpha) + s\omega^\alpha + o(\omega^\alpha).$$

$$= s\omega^\alpha + o(\omega^\alpha) \quad \text{as } \rho \text{ tends to infinity.}$$

This completes the proof of theorem.

Finally, an example shows that the theorem is a best possible theorem of its kind for the case $\beta = 1$.

NATIONAL UNIVERSITY OF CHEKIANG,
MEITAN, KWEICHOU,
CHINA.

REFERENCES

1. L. S. BOSANQUET, "*A solution of the Cesaro summability problem for successively derived Fourier series*," Proc. London Math. Soc., 2, vol. 46, 1940, 270–289.
2. L. S. BOSANQUET AND E. H. LINFOOT, "*Note on an asymptotic formula*," Tohoku Math. J., 39(1934), 11–16.
3. G. H. HARDY AND J. E. LITTLEWOOD, "*Solution of the Cesaro summability problem for power series and Fourier series*," Math. Zeit., 19(1924), 67–96.
4. J. J. GERGEN, "*Convergence and summability criteria for Fourier series*," Quarterly Math. Journ., 1(1930), 252–275.
5. F. T. WANG, "*On Riesz summability of Fourier series I, II*," Proc. London Math. Soc., vol. 47, 1942, 308–325.
6. E. C. TITCHMARSH, "*Theory of functions*," p. 107.

ANNALS OF MATHEMATICS
Vol. 44, No. 3, July, 1943

THE INVERSE NÖRLUND MEAN

BY FLORENCE M. MEARS

(Received November 16, 1942)

1. Introduction

Consider the series of complex terms, $\sum_{n=0}^{\infty} u_n$, and denote the corresponding sequence by $\{U_n\}$; then $U_n = \sum_{k=0}^{n} u_k$. When there is no possibility of confusion, we shall use the abbreviated notation $\sum u_n$ for $\sum_{n=0}^{\infty} u_n$. In this paper we are concerned exclusively with sequence to sequence transformations with triangular matrices, which assign to a given sequence $\{U_n\}$ the value $\lim_{n\to\infty} U'_n$, where

$$U'_n = \sum_{k=0}^{n} a_{nk} U_k, \tag{1}$$

provided that limit exists. If $U'_n \to U$, $\sum u_n$ is said to be summable to U' by the transformation (1); if $\sum | u'_n |$ converges, when $u'_n = U'_n - U'_{n-1}$, then $\sum u_n$ is said to be absolutely summable by (1).

We list for reference the following conditions:

$$\lim_{n\to\infty} a_{nk} = 0, \qquad k = 0, 1, \cdots; \tag{2}$$

$$\lim_{n\to\infty} \sum_{k=0}^{n} a_{nk} = C, \quad \text{where } C \text{ is a constant}; \tag{3}$$

$$\lim_{n\to\infty} \sum_{k=0}^{n} a_{nk} = 1; \tag{4}$$

$$\sum_{k=0}^{n} | a_{nk} | < M, \quad \text{where } M \text{ is a constant}; \qquad n = 0, 1, \cdots; \tag{5}$$

$$\left| \sum_{p=k}^{n} a_{np} \right| < M, \qquad n, k = 0, 1, \cdots; \tag{6}$$

$$\sum_{n=k}^{\infty} \left| \sum_{p=k}^{n} (a_{np} - a_{n-1,p}) \right| < M, \qquad k = 0, 1, \cdots. \tag{7}$$

The transformation (1) is said to be regular if $U'_n \to U$ whenever $U_n \to U$; it is said to be absolutely regular if $\sum | u'_n |$ converges whenever $\sum | u_n |$ converges. We recall the following theorems:

In order that (1) be regular, (2), (4) and (5) are necessary and sufficient.[1]

In order that $U'_n \to CU$ whenever $U_n \to U$, (2), (3) and (5) are necessary and sufficient.[2]

---

[1] Silverman, University of Missouri Studies, (1913), p. 49.
Toeplitz, Prace matematycznofizyczne, vol. 22 (1911), p. 117.

[2] Hausdorff, Mathematische Zeitschrift, vol. 9 (1921), p. 75.

In order that $U'_n \to CU$ whenever $\sum | u_n |$ converges, (2), (3) and (6) are necessary and sufficient.[3]

In order that (1) be absolutely regular, (7) is necessary and sufficient.[4]

In this paper we are concerned particularly with the Nörlund and Cesàro means, and with their inverses. The Nörlund mean, (N, p), is obtained by replacing a_{nk} of (1) by $P_n^{-1}p_{n-k}$, $n = 0, 1, \cdots, k = 0, 1 \cdots, n$, where $\{p_n\}$ is a given sequence of complex constants such that $P_n = \sum_{k=0}^n p_k \neq 0$.[5] Since $p_0 \neq 0$, and since the transformation is not affected if p_k is replaced by $p_0^{-1}p_k$, we shall let $p_0 = 1$. Then for (N, p), (1) becomes

$$U'_n = \sum_{k=0}^{n} P_n^{-1} p_{n-k} U_k, \qquad n = 0, 1, \cdots. \tag{8}$$

If $\sum u'_n$ converges, $\sum u_n$ is said to be summable (N, p); if $\sum | u'_n |$ converges, $\sum u_n$ is said to be summable $| N, p |$.

Let $$D_n^{(p)} = \begin{vmatrix} p_1 & 1 & 0\,0 \cdots 0 \\ p_2 & p_1 & 1\,0 \cdots 0 \\ \cdots & \cdots & \cdots \\ p_{n-1} & p_{n-2} & \cdots\cdots 1 \\ p_n & p_{n-1} & \cdots\cdots p_1 \end{vmatrix}, \qquad n = 1, 2, \cdots; \qquad D_0 = 1.$$

The inverse Nörlund mean, $(N, p)^{-1}$, is obtained by replacing a_{nk} of (1) by $(-1)^{n-k} P_k D_{n-k}^{(p)}$, $n = 0, 1, \cdots, k = 0, 1, \cdots, n$.[6] Therefore for $(N, p)^{-1}$, (1) becomes

$$U'_n = \sum_{k=0}^{n} (-1)^{n-k} P_k D_{n-k}^{(p)} U_k, \qquad n = 0, 1, \cdots. \tag{9}$$

If $\sum u'_n$ converges, we shall say that $\sum u_n$ is summable $(N, p)^{-1}$; if $\sum | u'_n |$ converges, we shall say that it is summable $| N, p |^{-1}$.

Let (N, p) be the Cesàro mean, (C, r), where r may take on any value except that of a negative integer; then $p_n = [n!\Gamma(r + 1)]^{-1} r\Gamma(r + n)$. If, for (C, r), we replace U'_n of (1) by $U_n^{(r)}$, we have

$$U_n^{(r)} = \frac{n!r}{\Gamma(r + n + 1)} \sum_{k=0}^{n} \frac{\Gamma(r + n - k)U_k}{(n - k)!}. \tag{10}$$

For the inverse Nörlund mean, $(N, p)^{-1}$, $a_{n,n-k} = P_{n-k}\, a_{k,0}$. If we let $(N, p)^{-1} = (C, r)^{-1}$, we can easily prove that $a_{k,0} = (-1)^k r(r - 1) \cdots (r - k + 1)(k!)^{-1}$, $k = 1, 2, \cdots$; therefore when r is integral $a_{n,n-k} = 0$ if $k > r$. If, for $(C, r)^{-1}$, we replace U'_n of (1) by $\tilde{U}_n^{(r)}$, we have

---

[3] Hahn, Monatshefte für Mathematik und Physik, vol. 32 (1922), p. 29. This theorem, proved by Hahn for real numbers only, holds also for complex numbers.

[4] Mears, Annals of Mathematics, vol. 38, No. 3 (1937), p. 595.

[5] Nörlund, Lunds Universitet Årsskrift, (2), vol. 16 (1920), No. 3, p. 8.

[6] Hill, Duke Mathematical Journal, vol. 3, No. 4 (1937), p. 705.

$$(11) \qquad \bar{U}_n^{(r)} = \sum_{k=0}^{N} (-1)^k \frac{\Gamma(r+n-k+1)U_{n-k}}{\Gamma(r-k+1)(n-k)!k!}, \qquad n = k, k+1, \cdots,$$

where $N = r$ for r integral, if $r \leqq n$, and $N = n$ in all other cases.

It is known that (C, r) is regular if and only if $r = 0$, or if the real component of r is greater than zero. We shall prove the corresponding theorem for absolute regularity.

Theorem 1: *The Cesàro mean, (C, r), is absolutely regular if and only if $r = 0$ or the real component of r is greater than zero.*

Proof: For the Nörlund mean, (N, p), condition (7) becomes

$$(12) \qquad \sum_{n=0}^{\infty} | P_n P_{k+n}^{-1} - P_{n-1} P_{k+n-1}^{-1} | < M, \qquad k = 1, 2, \cdots; P_{-1} = 0.$$

If we let $(N, p) = (C, r)$, $r = R + i\rho$, and if

$$f(r) = k | r | \sum_{n=0}^{\infty} \left| \frac{(k+n-1)!\Gamma(r+n)}{n!\Gamma(r+n+k+1)} \right|,$$

the left side of (12) is equal to $f(r)$. Since (C, r) is absolutely regular for $r = R \geqq 0$,[7] $f(R) < M$, $k = 1, 2, \cdots, R \geqq 0$. But for $R \neq 0, f(r) \leqq | R^{-1} r | f(R)$; therefore (7) is satisfied and (C, r) is absolutely regular if $r = 0$, or if $R > 0$. Since

$$\lim_{n \to \infty} \frac{n! n^z}{z(z+1) \cdots (z+n)} = \Gamma(z), \qquad z \neq 0, -1, \cdots,$$

we have

$$0 < m' < \left| \frac{n! n^z}{\Gamma(z+n+1)} \right| < M', \qquad n = 0, 1, \cdots.$$

Therefore $f(r) \geqq k | r | m'(M')^{-1} \sum_{n=0}^{\infty} [n(k+n)]^{-1}$, for $R < 0$, or for $R = 0$, $r \neq 0$, and (7) is not satisfied for these values of r.

In section 2 we shall prove theorems concerning summability $(C, r)^{-1}$ and $| C, r |^{-1}$, when r is restricted to real values; section 3 includes multiplication theorems for the Nörlund mean and inverse mean; section 4 includes multiplication theorems for the Cesàro mean and inverse mean.

2. Cesàro Summability

Throughout this section, we shall let

$$(13) \qquad x_n^{(s)} = s^{-1} \sum_{k=0}^{N-1} (-1)^k \frac{\Gamma(s+n-k)u_{n-k}}{\Gamma(s-k)k!(n-k-1)!},$$

where $M = s$ for s integral, $s \leqq n$, and $M = n$ in all other cases.

[7] Kogbetliantz, Bulletin des sciences mathématiques, (2), vol. 49 (1925), p. 237.

We have

$$\tilde{U}_n^{(1)} = U_n + nu_n; \qquad \tilde{U}_n^{(2)} = \tilde{U}_n^{(1)} + 2^{-1}[(n+1)nu_n - n(n-1)u_{n-1}],$$

and in general

THEOREM 2: *If r is neither zero nor a negative integer, then*

$$\tilde{U}_n^{(r)} = \tilde{U}_n^{(r-1)} + x_n^{(r)}. \tag{14}$$

PROOF: Let

$$S_p = (-1)^p \frac{n\Gamma(r+n-p)}{\Gamma(r+1-p)(n-p)!p!}.$$

Then

$$\tilde{U}_n^{(r)} - \tilde{U}_n^{(r-1)} = \sum_{p=0}^{N} S_p U_{n-p} = \sum_{k=0}^{N-1} \sum_{p=0}^{k} S_p u_{n-k} + \sum_{p=0}^{N} S_p U_{n-N}, \tag{15}$$

where N is defined as in (11). It can be proved that

$$\sum_{p=0}^{k} S_p = (-1)^k \frac{\Gamma(r+n-k)}{r\Gamma(r-k)k!(n-k-1)!}, \qquad k = 0, 1, \cdots, N-1;$$

therefore $\sum_{p=0}^{N-1} S_p = -S_N$ and $\sum_{p=0}^{N} S_p = 0$. Substituting these results in (15), we obtain (14).

THEOREM 3: *If $\sum u_n$ is summable $(C, r+s)^{-1}$ to U, it is summable $(C, r)^{-1}$ to U, $s \geqq 0$, $r > -1$.*

PROOF: For $s \geqq 0$, $r > -1$, $(C, r+s)$ includes (C, r);[8] therefore $(C, r+s) = A(C, r)$, where A represents a regular matrix. Since (C, r) and $(C, r+s)$, and hence their inverses, are permutable, we have

$$\begin{aligned}(C, r)^{-1} &= (C, r+s)(C, r+s)^{-1}(C, r)^{-1} = A(C, r)(C, r+s)^{-1}(C, r)^{-1} \\ &= A(C, r)(C, r)^{-1}(C, r+s)^{-1} = A(C, r+s)^{-1}.\end{aligned}$$

Therefore $(C, r)^{-1}$ includes $(C, r+s)^{-1}$.

COROLLARY 1: *If $\sum u_n$ is summable $(C, r)^{-1}$, $r > 0$, it is summable $(C, r-1)^{-1}$, and $\lim_{n\to\infty} x_n^{(r)} = 0$.*

THEOREM 4: *If $\sum u_n$ is summable $|C, r+s|^{-1}$, it is summable $|C, r|^{-1}$, $s \geqq 0$, $r > -1$.*

PROOF: For $s \geqq 0$, $r > -1$, $|C, r+s|$ includes $|C, r|$;[9] therefore the matrix A, defined in Theorem 3, is absolutely regular. The rest of the proof is the same as that of Theorem 3.

THEOREM 5: *If $\lim_{n\to\infty} x_n^{(r)} = 0$, $r > 1$, $n = 1, 2, \cdots$, then $\lim_{n\to\infty} x_n^{(r-1)} = 0$.*

PROOF: Consider the triangular matrix $\| a_{mp} \|$, defined by $a_{mp} = 0$, when r is integral, $p > r-2$; $a_{mp} = r(m+1)!\Gamma(r+p)[(p+1)!\Gamma(r+m+1)]^{-1}$, in all other cases. It is easily proved that $x_{m+1}^{(r-1)} = \sum_{p=0}^{m} a_{mp} x_{p+1}^{(r)}$, $m = 0, 1, \cdots$.

---

[8] Knopp, Sitzungsberichte der Berliner Mathematischen Gesellschaft, vol. 7 (1907), p. 5.

[9] McFadden, Duke Mathematical Journal, vol. 9, No. 1 (1942), p. 173.

But

$$\sum_{p=0}^{m} | a_{mp} | = r(r-1)^{-1} - r(m+1)!\Gamma(r-1)[\Gamma(r+m+1)]^{-1}$$
$$< r(r-1)^{-1} + r\Gamma(r-1) < M,$$

and

$$\lim_{m\to\infty} a_{mp} = r\Gamma(r+p-1)(p!)^{-1}\cdot\lim_{m\to\infty} m![\Gamma(r+m)]^{-1} = 0.$$

Therefore since $\| a_{mp} \|$ satisfies (2) and (5), the theorem is proved.[10]

THEOREM 6: *If* $\sum u_n$ *is summable* $(C, r-1)^{-1}$, $r > 0$, *and if* $\lim_{n\to\infty} x_n^{(r)} = l$, *then* $l = 0$.

PROOF: If the hypothesis is satisfied, $\sum u_n$ is summable $(C, r)^{-1}$, as a result of Theorem 2. Theorem 6 then follows from Theorem 3, Cor. 1.

THEOREM 7: *The condition,* $\lim_{n\to\infty} x_n^{(r)} = 0$, *does not imply the convergence of* $\sum u_n$.

PROOF: Let $u_0 = 1$, and let

$$u_m = \frac{(m-1)!}{\Gamma(r+m)} \sum_{p=2}^{m+1} \frac{(r-1)\Gamma(r+m-p)}{\Gamma(r)(m-p+1)!\log p}, \qquad m = 1, 2, \cdots,$$

where $(r-1)\Gamma(r+m-p)$ is replaced by 1 when $r = 1$, $p = m+1$. Let

$$S_{kp} = 0, \qquad k \geqq r, \qquad r \text{ integral and } \leqq n,$$
$$S_{kp} = (-1)^k \frac{(r-1)\Gamma(r+p-k-1)}{\Gamma(r-k)k!(p-k)!\log(n-p+1)},$$

all other k. Then

$$x_n^{(r)} = [\Gamma(r+1)]^{-1} \sum_{k=0}^{N} \sum_{p=k}^{n-1} S_{kp} = [\Gamma(r+1)]^{-1} \sum_{p=0}^{n-1} \sum_{k=0}^{p} S_{kp}. \tag{16}$$

Letting $k = p = 0$ in (16), we find that the coefficient of $[\log(n+1)]^{-1}$ is $[\Gamma(r+1)]^{-1}$. For $1 \leqq x \leqq p$, we can easily prove that $\sum_{k=0}^{x-1} S_{kp} = -x(r+p-x-1)[p(r-1)]^{-1}S_{xp}$. Therefore $\sum_{k=0}^{p-1} S_{kp} = -S_{pp}$ and $\sum_{k=0}^{p} S_{kp} = 0$, $p \geqq 1$. It follows that $x_n^{(r)} = [\Gamma(r+1)\log(n+1)]^{-1}$; therefore the hypothesis of the theorem is satisfied. But

$$u_m > \frac{(m-1)!}{\Gamma(r+m)\log(m+1)} \sum_{p=2}^{m+1} \frac{(r-1)\Gamma(r+m-p)}{\Gamma(r)(m-p+1)!}$$
$$= [(r+m-1)\log(m+1)\Gamma(r)]^{-1}, \qquad m = 1, 2, \cdots;$$

therefore $\sum u_n$ diverges.

THEOREM 8: *If* $u_n = (-1)^k p_n$, $n_k \leqq n \leqq n_k + a_k$, $k = 0, 1, \cdots$, *where* $p_n \geqq 0$, $n_0 = 0$, $n_k = n_{k-1} + a_{k-1} + 1$; *and if*

$$\sum u_n \textit{ is summable } (C, r)^{-1}, \qquad r \geqq 2, \tag{17}$$

[10] Toeplitz, loc. cit., p. 115.

then $\sum u_n$ *is summable* $|C, 0|$, *provided there is a positive constant* a *for which at least one of the following conditions is satisfied:*

$$a_{2k} < a, \qquad k = 0, 1, \cdots, \tag{18}$$

$$a_{2k+1} < a, \qquad k = 0, 1, \cdots. \tag{19}$$

PROOF: Using (17) and the corollary of Theorem 3, we find that $x_n^{(1)}$ and $x_n^{(2)} \to 0$ as $n \to \infty$. Therefore $n(n-1)|u_n - u_{n-1}| \to 0$ as $n \to \infty$, and we can find a positive integer N such that the following conditions are satisfied when $n > N$:

$$p_n \leqq p_n + p_{n-1} = |u_n - u_{n-1}| < [n(n-1)]^{-1}, \qquad n = n_k;$$

$$p_n - p_{n-1} \leqq |p_n - p_{n-1}| = |u_n - u_{n-1}| < [n(n-1)]^{-1}, \qquad n \neq n_k.$$

Assuming (18) satisfied, and replacing $\sum_{m=0}^{a_k} p_{n_k+m}$ by P_{n_k}, $k = 0, 1, \cdots$, we have, when $n_{2l} > N$

$$\sum_{k=l}^{\infty} P_{n_{2k}} < \sum_{k=l}^{\infty} \sum_{m=0}^{a_{2k}} \frac{a_{2k} - m + 1}{(n_{2k} + m)(n_{2k} + m - 1)} < \sum_{k=l}^{\infty} \sum_{m=0}^{a_{2k}} \frac{a + 1}{(n_{2k} + m)(n_{2k} + m - 1)}.$$

Therefore $\sum_{k=0}^{\infty} P_{n_{2k}}$ converges, and since $\sum u_n$ converges, $\sum_{k=0}^{\infty} P_{n_{2k+1}}$ converges also. This completes the proof when (18) is satisfied; a similar proof holds for (19).

3. Multiplication Theorems for the Nörlund Mean

Let $\sum w_n$ be the Cauchy product of the series $\sum u_n$ and $\sum v_n$. Consider the Nörlund means, (N, a) and (N, b), defined by the sequences $\{a_n\}$ and $\{b_n\}$. Let $\{c_n\}$ be the sequence defined by the equations $b_n = \sum_{k=0}^{n} a_{n-k}c_k$; we assume throughout this section that $\{a_n\}$ and $\{b_n\}$ are fixed sequences for which $\sum_{k=0}^{n} c_k = C_n \neq 0$. When a series $\sum v_n$ is assigned, let b'_n and c'_n be defined, for $n = 0, 1, \cdots$, by $b'_n = \sum_{k=0}^{n} b_k v_{n-k}$ and $c'_n = \sum_{k=0}^{n} c_k v_{n-k}$; and let b'_{-1} and $c'_{-1} = 0$. We shall make use of the following conditions:

(20) $\sum v_n$ summable (N, b) to V;

(21) $\sum v_n$ summable $|N, b|$;

$$\lim_{n \to \infty} B_n^{-1} c'_{n-k} = 0, \qquad k = 0, 1, \cdots; \tag{22}$$

$$|B_n^{-1}| \sum_{p=0}^{n} |A_p c'_{n-p}| < M, \qquad n = 0, 1, \cdots; \tag{23}$$

$$\left| B_n^{-1} \sum_{p=k}^{n} A_p c'_{n-p} \right| < M, \qquad n, k = 0, 1, \cdots; \tag{24}$$

$$\sum_{n=k}^{\infty} \left| \sum_{p=0}^{n-k} [c'_p B_n^{-1} - c'_{p-1} B_{n-1}^{-1}] A_{n-p} \right| < M, \qquad k = 0, 1, \cdots; \tag{25}$$

$$\lim_{n \to \infty} B_n^{-1} \sum_{p=0}^{n-k} A_{n-p}^{-1} a_{n-p-k} b'_p = 0, \qquad k = 0, 1, \cdots; \tag{26}$$

$$(27)\qquad |B_n^{-1}| \sum_{k=0}^{n} \left| \sum_{p=0}^{n-k} A_{n-p}^{-1} a_{n-p-k} b'_p \right| < M, \qquad n = 0, 1, \cdots;$$

$$(28)\qquad \left| B_n^{-1} \sum_{p=0}^{n-k} A_{n-p}^{-1} A_{n-p-k} b'_p \right| < M, \qquad n, k = 0, 1, \cdots;$$

$$(29)\qquad \sum_{n=k}^{\infty} \left| \sum_{p=0}^{n-k} [b'_p B_n^{-1} - b'_{p-1} B_{n-1}^{-1}] A_{n-p-k} A_{n-p}^{-1} \right| < M, \qquad k = 0, 1, \cdots.$$

THEOREM 9: *In order that $\sum v_n$ may be such that $\sum w_n$ is summable (N, b) to UV whenever $\sum u_n$ is summable (N, a) to U, it is necessary and sufficient that $\sum v_n$ satisfy* (20), (22) *and* (23).

Let (N, a) be regular, and assume that $\sum_{n=0}^{\infty} n \mid x_n \mid < \infty$, where x_n is defined, for $n = 1, 2, \cdots$, by $\sum_{k=0}^{n} x_{n-k} A_k = 0$, and where $x_0 = 1$. This restricted form of Theorem 9 has been obtained by C. N. Moore[11] as a special case of a more general theorem.

THEOREM 10: *In order that $\sum v_n$ may be such that $\sum w_n$ is summable (N, b) to UV whenever $\sum u_n$ is summable (N, a) to U and in addition summable $\mid N, a \mid$, it is necessary and sufficient that $\sum v_n$ satisfy* (20), (22) *and* (24).

THEOREM 11: *In order that $\sum v_n$ may be such that $\sum w_n$ is summable $\mid N, b \mid$ whenever $\sum u_n$ is summable $\mid N, a \mid$, it is necessary and sufficient that $\sum v_n$ satisfy* (25).

PROOFS of Theorems 9, 10, and 11: The sequence $\{U'_n\}$, into which $\{U_n\}$ is transformed by (N, a) is obtained from (8) when p_n is replaced by a_r, $n = 0, 1, \cdots$. The sequence $\{W_n\}$ is transformed by (N, b) into $\{W'_n\}$, where

$$(30)\qquad W'_n = B_n^{-1} \sum_{k=0}^{n} v_{n-k} \sum_{p=0}^{k} b_{k-p} U_p .$$

The matrix $\| c_{nk} \|$, defined by $c_{nk} = B_n^{-1} A_k c'_{n-k}$, transforms $\{U'_n\}$ into $\{W'_n\}$. To complete the proof of Theorem 9, we require that $\| c_{nk} \|$ satisfy (2), (3) and (5); to complete Theorem 10, that $\| c_{nk} \|$ satisfy (2), (3) and (6); to complete Theorem 11, that $\| c_{nk} \|$ satisfy (7).

THEOREM 12: *In order that $\sum v_n$ may be such that $\sum w_n$ is summable (N, b) to UV whenever $\sum u_n$ is summable $(N, a)^{-1}$ to U, it is necessary and sufficient that $\sum v_n$ satisfy* (20), (26) *and* (27).

THEOREM 13: *In order that $\sum v_n$ may be such that $\sum w_n$ is summable (N, b) to UV whenever $\sum u_n$ is summable $(N, a)^{-1}$ to U and in addition summable $\mid N, a \mid^{-1}$, it is necessary and sufficient that $\sum v_n$ satisfy* (20), (26) *and* (28).

THEOREM 14: *In order that $\sum v_n$ may be such that $\sum w_n$ is summable $\mid N, b \mid$ whenever $\sum u_n$ is summable $\mid N, a \mid^{-1}$, it is necessary and sufficient that $\sum v_n$ satisfy* (29).

PROOFS of Theorems 12, 13 and 14: The sequence $\{U'_n\}$, into which $\{U_n\}$ is transformed by $(N, a)^{-1}$, is obtained from (9) when p_n is replaced by a_n,

---

[11] Moore, Summable Series and Convergence Factors, American Mathematical Society, Colloquium Publication, vol. 22, p. 44, Theorem II.

$n = 0, 1, \cdots$; $\{W'_n\}$ is defined by (30). The matrix $\| c_{nk} \|$, defined by $c_{nk} = B_n^{-1}[\sum_{p=k}^{n} A_p^{-1} a_{p-k} b'_{n-p}]$, transforms $\{U'_n\}$ into $\{W'_n\}$. The remainder of the proof is the same as that of Theorems 9, 10 and 11.

It is easy to prove that necessary and sufficient conditions that (22) follow from (20) when $(N, a) = (C, 0)$ are that $b_n B_{n+k}^{-1} \to 0$ and $| B_n B_{n+k}^{-1} | < M$, $n, k = 0, 1, \cdots$; a necessary and sufficient condition that (24) follow from (20) when $(N, a) = (C, 0)$ is $| B_n B_{n+k}^{-1} | < M$, $n, k = 0, 1, \cdots$; a necessary and sufficient condition that (25) follow from (21) when $(N, a) = (C, 0)$ is that (N, b) be absolutely regular. Using these facts we can prove the following theorems, which are special cases of Theorems 10 and 11.

THEOREM 10′: *Let (N, b) be regular. In order that $\sum v_n$ may be such that $\sum w_n$ is summable (N, b) to UV whenever $\sum u_n$ is summable $(C, 0)$ to U and in addition summable $| C, 0 |$, it is necessary and sufficient that $\sum v_n$ satisfy* (20).

THEOREM 11′: *Let (N, b) be absolutely regular. In order that $\sum v_n$ may be such that $\sum w_n$ is summable $| N, b |$ whenever $\sum u_n$ is summable $| C, 0 |$ it is necessary and sufficient that $\sum v_n$ satisfy* (21).

The triangular matrix $\| c_{np} \|$, defined by $c_{np} = B_p B_{n+k}^{-1}[A_{n-p} A_{n-p+k}^{-1} - A_{n-p-1} A_{n-p+k-1}^{-1}]$, $p = 0, 1, \cdots, n$; $A_{-1} = 0$, transforms $\{B_n^{-1} \sum_{p=0}^{n} B_{n-p} v_p\}$ into $\{B_{n+k}^{-1} \sum_{p=0}^{n} A_{n-p+k}^{-1} A_{n-p} b'_p\}$. It can be proved that necessary and sufficient conditions that $\| c_{np} \|$ transform every convergent sequence into a bounded sequence are that (N, a) be absolutely regular, and that $| B_n B_{n+k}^{-1} | < M$, $n, k = 0, 1, \cdots$; therefore, under these conditions, (28) follows from (20). Since $\| c_{np} \|$ is absolutely regular if (N, a) and (N, b) are absolutely regular, (29) is a consequence of (21) if these conditions on (N, a) and (N, b) are satisfied. Similarly we can prove that (26) is a consequence of (20) if (N, a) and (N, b) are regular and (N, a) absolutely regular. From these facts we obtain the following special cases of Theorems 13 and 14.

THEOREM 13′: *Let (N, a) and (N, b) be regular, and (N, a) absolutely regular. In order that $\sum v_n$ may be such that $\sum w_n$ is summable (N, b) to UV whenever $\sum u_n$ is summable $(N, a)^{-1}$ to U and in addition summable $| N, a |^{-1}$, it is necessary and sufficient that $\sum v_n$ satisfy* (20).

THEOREM 14′: *Let (N, a) and (N, b) be absolutely regular. In order that $\sum v_n$ may be such that $\sum w_n$ is summable $| N, b |$ whenever $\sum u_n$ is summable $| N, a |^{-1}$ it is necessary and sufficient that $\sum v_n$ satisfy* (21).

Comparing Theorems 10′ and 13′, we find that in order that $\sum w_n$ may be summable (N, b) to UV when (N, a) and (N, b) are regular and (N, a) absolutely regular, it is necessary and sufficient that $\sum v_n$ satisfy (20), not only if $\sum u_n$ is summable $| C, 0 |$, but also if $\sum u_n$ satisfies the more stringent condition of summability $| N, a |^{-1}$. A similar comparison of Theorems 11′ and 14′ shows that in order that $\sum w_n$ may be summable $| N, b |$, when (N, a) and (N, b) are absolutely regular, it is necessary and sufficient that $\sum v_n$ satisfy (21), whether $\sum u_n$ be summable $| N, a |^{-1}$, or merely summable $| C, 0 |$.

However, if we do not impose additional conditions upon (N, a) and (N, b), we can construct examples to show that the conditions imposed upon $\sum v_n$

in Theorem 13 are less stringent than those of Theorem 10, and those of Theorem 14 less stringent than those of Theorem 11. For if we let $(N, b) = (C, -1/2)$, $v_0 = 1, v_n = 0, n = 1, 2, \cdots$, we find that when $(N, a)^{-1} = (C, 1)^{-1}$, (20), (26), (28) and (29) are satisfied, but when $(N, a) = (C, 0)$, (24) and (25) are not satisfied. Therefore there is some $\sum u_n$ summable $| C, 0 |$, such that $\sum w_n = \sum u_n$ is not summable $(C, -1/2)$ [or $| C, -1/2 |$], but there is no $\sum u_n$ summable $| C, 1 |^{-1}$ such that $\sum w_n$ is not summable $(C, -1/2)$ [or $| C, -1/2 |$].

Conditions of regularity and absolute regularity of (N, a) and (N, b) are not sufficient to make the conditions of Theorem 12 equivalent to those of Theorem 9 when $(N, a) = (C, 0)$. If $(N, b) = (C, 0)$, $\sum_{n=0}^{\infty} (-1)^n(n + 1)^{-1}$ satisfies the conditions of Theorem 12 when $(N, a) = (C, 1)$, but does not satisfy Theorem 9 when $(N, a) = (C, 0)$.

4. Multiplication Theorems for the Cesàro Mean

If we let $(N, a) = (C, r)$ and $(N, b) = (C, s)$, where r and s are either equal to zero or have a real component greater than zero, (C, r) and (C, s) are regular and absolutely regular. It is easily proved that in this case (22) is a consequence of (20). Therefore we shall not need conditions corresponding to (22), (26), (28) and (29) for the Cesàro mean. In the following conditions, $x\Gamma(x + y - 1)$ is to be replaced by 1 when $x = 0$, $y = 1$; $s \geqq r$ in (33), (34) and (35); and $S(x, y) = \sum_{q=0}^{y} x\Gamma(x + q)[q!\Gamma(x + 1)]^{-1}v_{y-q}$, $y = 0, 1, \cdots$; $S(x, -1) = 0$.

(31) $\sum v_n$ summable (C, s) to V;

(32) $\sum v_n$ summable $| C, s |$;

$$(33)\quad \sum_{p=0}^{n} \left| \frac{n!\Gamma(r + p + 1)S(s - r, n - p)}{\Gamma(s + n + 1)p!} \right| < M, \qquad n = 0, 1, \cdots;$$

$$(34)\quad \left| \sum_{p=k}^{n} \frac{n!\Gamma(r + p + 1)S(s - r, n - p)}{\Gamma(s + n + 1)p!} \right| < M, \qquad n, k = 0, 1, \cdots;$$

$$(35)\quad \sum_{n=k}^{\infty} \left| \sum_{p=0}^{n-k} \left[\frac{n!S(s - r, p)}{\Gamma(s + n + 1)} - \frac{(n - 1)!S(s - r, p - 1)}{\Gamma(S + n)} \right] \frac{\Gamma(r + n - p + 1)}{(n - p)!} \right| < M, \quad k = 0, 1, \cdots;$$

$$(36)\quad \sum_{k=0}^{n} \left| \sum_{p=0}^{n-k} \frac{n!r\Gamma(r + n - p - k)(n - p)!S(s, p)}{\Gamma(s + n + 1)(n - p - k)!\Gamma(r + n - p + 1)} \right| < M, \quad n = 0, 1, \cdots;$$

The following theorems are special cases of Theorems 9 through 14.

THEOREM 15: *In order that $\sum v_n$ may be such that $\sum w_n$ is summable (C, s) to UV whenever $\sum u_n$ is summable (C, r) to U, it is necessary and sufficient that $\sum v_n$ satisfy* (31) *and* (33).[12]

THEOREM 16: *In order that $\sum v_n$ may be such that $\sum w_n$ is summable (C, s)*

[12] Moore, loc. cit., p. 46, Theorem VI.

to UV whenever $\sum u_n$ *is summable* (C, r) *to* U *and in addition summable* $|C, r|$, *it is necessary and sufficient that* $\sum v_n$ *satisfy* (31) *and* (34).

THEOREM 17: *In order that* $\sum v_n$ *may be such that* $\sum w_n$ *is summable* $|C, s|$ *whenever* $\sum u_n$ *is summable* $|C, r|$, *it is necessary and sufficient that* $\sum v_n$ *satisfy* (35).

THEOREM 18: *In order that* $\sum v_n$ *may be such that* $\sum w_n$ *is summable* (C, s) *to UV whenever* $\sum u_n$ *is summable* $(C, r)^{-1}$ *to* U, *it is necessary and sufficient that* $\sum v_n$ *satisfy* (31) *and* (36).

THEOREM 19: *In order that* $\sum v_n$ *may be such that* $\sum w_n$ *is summable* (C, s) *to UV whenever* $\sum u_n$ *is summable* $(C, r)^{-1}$ *to* U *and in addition summable* $|C, r|^{-1}$, *it is necessary and sufficient that* $\sum v_n$ *satisfy* (31).

THEOREM 20: *In order that* $\sum v_n$ *may be such that* $\sum w_n$ *is summable* $|C, s|$ *whenever* $\sum u_n$ *is summable* $|C, r|^{-1}$, *it is necessary and sufficient that* $\sum v_n$ *satisfy* (32).

THE GEORGE WASHINGTON UNIVERSITY.

ANNALS OF MATHEMATICS
Vol. 44, No. 3, July, 1943

CONVERGENCE OF CERTAIN GAP SERIES

BY M. KAC

(Received March 2, 1942)

1. The present paper investigates convergence properties of series

$$\sum_{k=1}^{\infty} c_k \varphi(n_k x),$$

where $\{n_k\}$ is a gap sequence of integers ($n_{k+1}/n_k > q > 1$) and $\varphi(x)$ a periodic function satisfying Hölder's condition. Several results in this direction were obtained by the present writer[1] and Izumi and Kawata.[2] In both cases not only bigger gaps were considered but also the sequence $\{n_k\}$ was assumed to be of the form 2^{m_k}, where $\{m_k\}$ was a further specified sequence of integers. In this paper nothing is assumed about the arithmetical structure of the sequence $\{n_k\}$ and the gap condition is the one to be hoped for by analogy with a corresponding theorem of Kolmogoroff[3] concerning trigonometrical series. In concluding the proof we follow a very ingenious idea of Paley and Zygmund.[4]

2. THEOREM. *Let $\varphi(x)$ be a complex-valued function, defined over $-\infty < x < \infty$, satisfying the conditions*

$$\varphi(x + \lambda) = \varphi(x) \qquad -\infty < x < \infty \tag{1}$$

$$| \varphi(x') - \varphi(x'') | \leqq M | x' - x'' |^{\alpha} \qquad -\infty < x', x'' < \infty \tag{2}$$

$$\int_0^{\lambda} \varphi(x)\, dx = 0 \tag{3}$$

where λ, M, and α are constants for which $\lambda > 0$ and $0 < \alpha \leqq 1$. Let $\{n_k\}$ be a sequence of integers satisfying the gap condition $n_{k+1}/n_k > q > 1$, $k = 1, 2, \cdots$, for some constant q. Then if $\sum_{k=1}^{\infty} | c_k |^2 < \infty$, the series

$$\sum_{k=1}^{\infty} c_k \varphi(n_k x) \tag{4}$$

converges in the mean with exponent 2 over each finite interval to a function $f(x)$ having period λ and, moreover, the series (4) converges almost everywhere to this function $f(x)$.

---

[1] M. Kac. Sur les fonctions indépendantes V, Studia Mathematica, 7 (1938), pp. 96–100.

[2] S. Izumi and T. Kawata, On certain series of functions, Tohoku Math. Journal, 46 (1939), pp. 91–105.

[3] A. Kolmogoroff, Une contribution a l'étude de la convergence des séries de Fourier, Fundamenta Math. 5 (1924), pp. 96–97.

[4] See for instance S. Kaczmarz and H. Steinhaus, Theorie der Orthogonelreihen, Monografje Matematyczne, pp. 126–127 and 137–138, where the method is explained in two examples.

A linear change of independent variable reduces the general case to that in which $\lambda = 2\pi$; we hereafter take λ to be 2π. If $\varphi(x) = \varphi_1(x) + i\varphi_2(x)$ where $\varphi_1(x)$ and $\varphi_2(x)$ are real, then $\varphi_1(x)$ and $\varphi_2(x)$ each satisfy the conditions imposed upon $\varphi(x)$; it is therefore easy to see that the theorem follows when we prove it for the case in which $\varphi(x)$ is real. Likewise, it is sufficient to establish the result for the case in which the constants c_k are real. If the conditions on $\varphi(x)$ are satisfied when $\alpha = \alpha_2$ and if $0 < \alpha_1 < \alpha_2$, then the conditions are satisfied when $\alpha = \alpha_1$. Hence we may (and shall) assume that $0 < \alpha < 1$.

LEMMA 1. *There is a constant A independent of j and k such that*

$$\left| \int_0^{2\pi} \varphi(n_j x)\varphi(n_k x)\, dx \right| \leqq A(q^{-\alpha|j-k|}) \qquad j, k = 1, 2, \cdots.$$

Let $\varphi(x) \sim \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$, let $s_n(x)$ denote the n^{th} partial sum of this series, and let $\sigma_n(x)$ denote the n^{th} Fejér trigonometric polynomial of $\varphi(x)$. By a theorem of *S.* Bernstein,[5] there is a constant D such that

$$|\varphi(x) - \sigma_n(x)| \leqq Dn^{-\alpha} \qquad 0 \leqq x \leqq 2\pi,\ n = 1, 2, \cdots;$$

hence

$$\sum_{l=n+1}^{\infty} (a_l^2 + b_l^2) = \frac{1}{\pi} \int_0^{2\pi} |\varphi(x) - s_n(x)|^2\, dx$$

$$\leqq \frac{1}{\pi} \int_0^{2\pi} |\varphi(x) - \sigma_n(x)|^2\, dx \leqq Dn^{-2\alpha}.$$

Use of Parseval's relation gives

$$\frac{1}{\pi} \int_0^{2\pi} \varphi(n_j x)\varphi(n_k x)\, dx = \sum_{rn_j = sn_k} (a_r a_s + b_r b_s).$$

Assuming that $j < k$, we see that if positive integers r and s satisfy the equality $rn_j = sn_k$, then $s \geqq 1$ and $r \geqq (n_k/n_j)$; hence use of the inequality

$$|a_r a_s + b_r b_s| \leqq (a_r^2 + b_r^2)^{\frac{1}{2}}(a_s^2 + b_s^2)^{\frac{1}{2}}$$

and of the Schwarz inequality gives

$$\left| \frac{1}{\pi} \int_0^{2\pi} \varphi(n_j x)\varphi(n_k x)\, dx \right| \leqq \left[\sum_{r \geqq n_k/n_j} (a_r^2 + b_r^2) \right]^{\frac{1}{2}} \left[\sum_{s=1}^{\infty} (a_s^2 + b_s^2) \right]^{\frac{1}{2}}$$

$$\leq B(n_k/n_j)^{-\alpha} \leq B(q^{-\alpha|k-j|}).$$

This establishes Lemma 1.

LEMMA 2. *There is a constant C, independent of a, b, and k, such that*

$$\left| \int_a^b \varphi(n_k x)\, dx \right| \leqq Cn_k^{-1} \qquad 0 \leqq a < b \leqq 2\pi,\ k = 1, 2, \cdots.$$

[5] S. Bernstein, Sur l'ordre de la meilleure approximation des fonctions continues par des polynomes de degré donné, Memoires de l'Académie royal de Belgique, IV, pp. 1–104; in particular p. 88. See also A. Zygmund, Trigonometrical series, Monografje Matematyczne, p. 62.

When $0 \leqq x < 2\pi$, let $\psi(x)$ be the characteristic function of the interval $a \leqq x \leqq b$; and let $\psi(x)$ be extended over the infinite interval so that $\psi(x + 2\pi) = \psi(x)$. Let

$$\psi(x) \sim \tfrac{1}{2}d_0 + \sum_{n=1}^{\infty} (d_n \cos nx + e_n \sin nx).$$

Then, by Parseval's relation,

$$\frac{1}{\pi}\int_0^{2\pi} \psi(x)\varphi(n_k x)\, dx = \sum_{r=1}^{\infty} (a_r d_{rn_k} + b_r e_{rn_k})$$

and therefore

$$\left|\frac{1}{\pi}\int_0^{2\pi} \psi(x)\varphi(n_k x)\, dx\right| \leqq \left[\sum_{r=1}^{\infty} (a_r^2 + b_r^2)\right]^{\frac{1}{2}}\left[\sum_{r=1}^{\infty} (d_{rn_k}^2 + e_{rn_k}^2)\right]^{\frac{1}{2}}$$

$$\leqq \left[\sum_{r=1}^{\infty} (a_r^2 + b_r^2)\right]^{\frac{1}{2}}\left[\sum_{s=1}^{\infty} \frac{1}{s^2}\right]^{\frac{1}{2}} 2\sqrt{2}\, n_k^{-1}.$$

The conclusion of Lemma 2 follows.

LEMMA 3. *The series*

$$\sum_{k=1}^{\infty} c_k\varphi(n_k x)$$

converges in the mean, with exponent 2, *over* $0 \leqq x \leqq 2\pi$.

If m and n are integers for which $1 \leqq m < n$, then

$$\int_0^{2\pi} \left(\sum_{k=m}^{n} c_k\varphi(n_k x)\right)^2 dx = \sum_{j,k=m}^{n} c_j c_k \int_0^{2\pi} \varphi(n_j x)\varphi(n_k x)\, dx$$

$$\leqq A \sum_{j,k=m}^{n} \frac{|c_j|\,|c_k|}{q^{\alpha|j-k|}} \leqq 2A \sum_{r=0}^{n-m} \sum_{s=m+r}^{n} \frac{|c_s|\,|c_{s-r}|}{q^{\alpha r}}$$

$$\leqq 2A \sum_{r=0}^{n-m} \frac{1}{q^{\alpha r}} \sum_{s=m}^{n} c_s^2 \leqq 2A \sum_{r=0}^{\infty} \frac{1}{q^{\alpha r}} \sum_{s=m}^{n} c_s^2$$

and Lemma 3 follows.

Let $f(x)$ be the function to which $\sum c_k\varphi(n_k x)$ converges in the mean; this function has period 2π and belongs to class L_2 over the interval $0 \leqq x \leqq 2\pi$. Let t_0 be such a point that

$$f(t_0) = \lim_{h\to 0} \frac{1}{h}\int_{t_0}^{t_0+h} f(x)\, dx;$$

almost all points t have this property. We complete the proof of the theorem by proving that the series (4) converges to $f(t_0)$ when $x = t_0$.

Since $\sum c_k\varphi(n_k x)$ converges in mean to $f(x)$,

$$\int_{t_0}^{t_0+h} f(x)\, dx = \sum_{k=1}^{\infty} c_k \int_{t_0}^{t_0+h} \varphi(n_k x)\, dx;$$

using this fact and Lemma 2, we obtain

$$\left| \int_{t_0}^{t_0+h} f(x)\,dx - \sum_{k=1}^{r} c_k \int_{t_0}^{t_0+h} \varphi(n_k x)\,dx \right|$$
$$\leqq \left[\sum_{k=r+1}^{\infty} c_k^2\right]^{\frac{1}{2}} \left[\sum_{k=r+1}^{\infty} \frac{C^2}{n_k^2}\right]^{\frac{1}{2}} \leqq C_1 \left[\sum_{k=r+1}^{\infty} c_k^2\right]^{\frac{1}{2}} n_r^{-1}.$$

Use of (2) gives

$$\left| \sum_{k=1}^{r} c_k \int_{t_0}^{t_0+h} [\varphi(n_k x) - \varphi(n_k t_0)]\,dx \right| \leqq M |h|^{1+\alpha} \sum_{k=1}^{r} |c_k| n_k^{\alpha}.$$

Combining these inequalities gives

$$\left| \frac{1}{h} \int_{t_0}^{t_0+h} f(x)\,dx - \sum_{k=1}^{r} c_k \varphi(n_k t_0) \right| \leqq C_1 |h|^{-1} \left[\sum_{k=r+1}^{\infty} c_k^2\right]^{\frac{1}{2}} n_r^{-1} + M |h|^{\alpha} \sum_{k=1}^{r} |c_k| n_k^{\alpha}.$$

Let $h_r = n_r^{-1}$. Then

$$|h_r|^{\alpha} \sum_{k=1}^{r} |c_k| n_k^{\alpha} = \sum_{k=1}^{r} |c_k| (n_k/n_r)^{\alpha} \leqq \sum_{k=1}^{r} |c_k| q^{-\alpha(r-k)} = \frac{1}{q^{\alpha r}} \sum_{k=1}^{r} |c_k| q^{\alpha k};$$

and since $c_k \to 0$ as $k \to \infty$, the elementary theory of matrix transformations implies that the last member converges to 0 as $r \to \infty$. It now follows easily that

$$\lim_{r\to\infty} \left| \frac{1}{h_r} \int_{t_0}^{t_0+h_r} f(x)\,dx - \sum_{k=1}^{r} c_k \varphi(n_k t_0) \right| = 0$$

and hence that $\sum c_k \varphi(n_k t_0)$ converges to $f(t_0)$. This completes the proof of the theorem.

3. By analogy with some former results (loc. cit. 1 and 2) one might expect that under the conditions of our theorem the divergence of

$$\sum_{k=1}^{\infty} c_k^2 \qquad (c_n \to 0 \text{ as } n \to \infty)$$

would imply the divergence almost everywhere of the series

$$\text{(a)} \qquad \sum_{k=1}^{\infty} c_k \varphi(n_k x).$$

This can be easily disproved by the following simple example.

Let $\psi(x)$ satisfy the conditions (1), (2) and (3) of our theorem and put

$$\varphi(x) = \psi(x) - \psi(2x).$$

Put $n_k = 2^k$ and choose $\{c_k\}$ in such a way that

$$\text{(b)} \qquad \sum_{k=1}^{\infty} c_k^2 = \infty$$

and

$$\text{(c)} \qquad \sum_{k=1}^{\infty} | c_{k+1} - c_k | < \infty$$

(for instance $c_k = 1/\sqrt{k}$). We have

$$\sum_{k=1}^{n} c_k \varphi(2^k x) = c_1 \psi(2x) + \sum_{k=1}^{n-1} (c_{k+1} - c_k)\psi(2^{k+1} x) - c_n \psi(2^{n+1} x).$$

The right side tends to a limit for every x, hence so does the series (a). This combined with (b) disproves the conjecture.

CORNELL UNIVERSITY.

Annals of Mathematics
Vol. 44, No. 3, July, 1943

ON THE EXPANSION OF THE PARTITION FUNCTION IN A SERIES

By Hans Rademacher

(Received April 8, 1943)

1. A geometric property of the Farey series, discovered by L. R. Ford (**1**) is used in this note for the construction of a new path of integration to replace the circle carrying the Farey dissection, first introduced by Hardy and Ramanujan in their classical paper (**2**). This new path of integration will bring about an essential simplification in the treatment of the partition function and, in general, in the determination of the coefficients of modular functions of non-negative dimension. It seems to me that the new path exhibits more clearly than the Farey arcs do the different contributions of the approximation functions near the roots of unity. Moreover, only two estimations have to be performed, and they are direct consequences of the obvious statements (3.2) and (4.1) concerning the circle over the diameter 0 to 1.

Ford's theorem referred to above can be enunciated as follows:

If in a complex τ-plane we mark the points corresponding to the reduced fractions h/k and draw about the points

$$\tau_{h,k} = \frac{h}{k} + \frac{i}{2k^2}$$

the circles of radii $1/2k^2$, which touch the real axis at h/k, then these circles do not intersect. Two of them are tangent to each other if and only if their fractions h/k and l/m appear as neighbors in a Farey series of some order.

The proof is clear. Comparing the distance of the centers with the sum of the radii of two such circles we consider

$$|\tau_{h,k} - \tau_{l,m}|^2 - \left(\frac{1}{2k^2} + \frac{1}{2m^2}\right)^2 = \frac{(hm - lk)^2 - 1}{k^2m^2} \geqq 0.$$

The equality sign, indicating contact of the circles, is attained only for

$$hm - lk = \pm 1,$$

and this would mean that the fractions h/k and l/m are neighbors in some Farey series, e.g. that of order $N = k + m - 1$.

To each positive integer N we introduce now a path P_N which we shall later use for the complex integration. Let $c_{h,k}$ be the circle of Ford's theorem belonging to the (reduced) fraction h/k. We draw all circles $c_{h,k}$ for $k \leqq N$, $0 \leqq h/k < 1$, in other words all the circles belonging to the Farey series of order N. If $h_1/k_1 < h/k < h_2/k_2$ are three adjacent fractions of that series, then the circle $c_{h,k}$ has a point of contact with c_{h_1,k_1} as well as with c_{h_2,k_2}. These points of contact cut $c_{h,k}$ into two arcs, an upper one and a lower one. (The lower one touches the real axis.) As path P_N we take now the row of upper arcs $\gamma_{h,k}$, each traversed on its circle in the negative sense, on $c_{h,k}$ therefore from

the point of contact with c_{h_1,k_1} to the point of contact with c_{h_2,k_2}. The figure shows P_N for $N = 3$. Because of the periodicity of the function to be integrated it does not matter that instead of the whole arc $\gamma_{0,1}$ we have drawn a part of $\gamma_{1,1}$ which is obtained from the omitted part of $\gamma_{0,1}$ through the translation $+1$.

For a later purpose we need the coordinates of the endpoints of the arc $\gamma_{h,k}$. They are, as simple geometric arguments show,

$$\frac{h}{k} + \left(-\frac{k_1}{k(k^2+k_1^2)} + \frac{i}{k^2+k_1^2}\right) = \frac{h}{k} + \zeta'_{h,k} \tag{1.1}$$

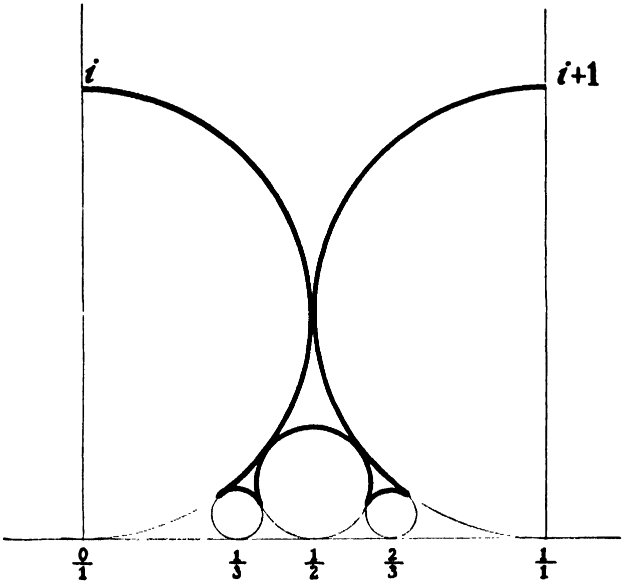


FIG. 1

and

$$\frac{h}{k} + \left(\frac{k_2}{k(k^2+k_2^2)} + \frac{i}{k^2+k_2^2}\right) = \frac{h}{k} + \zeta''_{h,k}. \tag{1.2}$$

Incidentally, the point $h/k + \zeta'_{h,k}$ lies on the semicircle over the diameter $(h_1/k_1, h/k)$; the whole path P_N lies above the row of semicircles connecting adjacent fractions of the Farey series of order N.[1]

[1] The imaginary part of $\zeta'_{h,k}$ is $1/(k^2 + k_1^2)$, and we have

$$\frac{1}{2N^2} \leq \frac{1}{k^2+k_1^2} \leq \frac{2}{(k+k_1)^2} \leq \frac{2}{(N+1)^2}.$$

Thus $\mathfrak{I}(\zeta'_{h,k})$ is of the order N^{-2}. This corresponds to the choice of a circle of radius $\exp(-2\pi N^{-2})$ as the path of integration associated with a Farey dissection of order N in my previous treatment of $p(n)$. (3).

2. In order to come to $p(n)$ we start with Euler's formula

$$(2.1) \qquad 1 + \sum_{n=1}^{\infty} p(n)x^n = \prod_{m=1}^{\infty} (1 - x^m)^{-1} = f(x), \qquad |x| < 1.$$

We have therefore

$$p(n) = \int_i^{i+1} f(e^{2\pi i\tau})e^{-2\pi in\tau}\, d\tau$$

for any path of integration in the upper τ-halfplane connecting i and $i + 1$. We choose the path P_N described above and obtain

$$p(n) = \sum_{0 \leq h < k \leq N} \int_{\gamma_{h,k}} f(e^{2\pi i\tau})e^{-2\pi in\tau}\, d\tau$$

$$= \sum_{0 \leq h < k \leq N} \int_{\zeta'_{h,k}}^{\zeta''_{h,k}} f(e^{2\pi i\left(\frac{h}{k}+\zeta\right)})e^{-2\pi in\left(\frac{h}{k}+\zeta\right)}\, d\zeta.$$

Here and in the sequel it is always understood that h is prime to k. The path of ζ between $\zeta'_{h,k}$ and $\zeta''_{h,k}$ is described by the substitution

$$\tau = \frac{h}{k} + \zeta,$$

where τ runs on $\gamma_{h,k}$. That means that ζ runs from $\zeta'_{h,k}$ to $\zeta''_{h,k}$ on the upper arc of a circle of radius $1/2k^2$ about the point $i/2k^2$. Introducing in each integral a new variable z by

$$\zeta = \frac{iz}{k^2}$$

we obtain

$$(2.2) \qquad p(n) = \sum_{0 \leq h < k \leq N} \frac{i}{k^2} e^{-\frac{2\pi inh}{k}} \int_{z'_{h,k}}^{z''_{h,k}} f(e^{\frac{2\pi ih}{k} - \frac{2\pi z}{k^2}})e^{\frac{2\pi nz}{k^2}}\, dz.$$

Here z runs in each integral on an arc of the circle K of radius $\frac{1}{2}$ about the point $\frac{1}{2}$ as center. The ends of the arc are

$$z'_{h,k} = -ik^2\zeta'_{h,k}, \qquad z''_{h,k} = -ik^2\zeta''_{h,k}$$

or, according to (1.1) and (1.2)

$$(2.3) \qquad \begin{aligned} z'_{h,k} &= \frac{k^2}{k^2 + k_1^2} + i\,\frac{kk_1}{k^2 + k_1^2}, \\ z''_{h,k} &= \frac{k^2}{k^2 + k_2^2} - i\,\frac{kk^2}{k^2 + k_2^2}. \end{aligned}$$

The points $z'_{h,k}$ and $z''_{h,k}$ divide the circle K into two arcs; that one is meant as path of integration which does not touch the imaginary axis. We have $\Re(z) > 0$.

On the integrands in (2.2) we apply now the transformation formula of the function $f(x)$, a formula, which stems from the theory of the elliptic modular functions (cf. (**2**), Lemma 4.32)

$$f\left(e^{\frac{2\pi ih}{k}-\frac{2\pi z}{k^2}}\right) = \omega_{h,k}\sqrt{\frac{z}{k}}\exp\left(\frac{\pi}{12z}-\frac{\pi z}{12k^2}\right)f\left(e^{\frac{2\pi ih'}{k}-\frac{2\pi}{z}}\right).$$

Here $\omega_{h,k}$ is a well-known root of unity, and h' is a solution of the congruence

$$hh' \equiv -1 \pmod{k};$$

for the square root the principal branch has to be taken. We get therefore from (2.2)

$$p(n) = \sum_{0\leq h<k\leq N} ik^{-\frac{1}{2}}\omega_{h,k}e^{-\frac{2\pi ihn}{k}}\int_{z'_{h,k}}^{z''_{h,k}}\Psi_k(z)f\left(e^{\frac{2\pi ih'}{k}-\frac{2\pi}{z}}\right)e^{\frac{2\pi nz}{k^2}}\,dz, \tag{2.4}$$

with the abbreviation

$$\Psi_k(z) = z^{\frac{1}{2}}\exp\left(\frac{\pi}{12z}-\frac{\pi z}{12k^2}\right). \tag{2.5}$$

We rewrite (2.4) as

$$\begin{aligned} p(n) &= \sum_{0\leq h<k\leq N} ik^{-\frac{1}{2}}\omega_{h,k}e^{-\frac{2\pi ihn}{k}}\int_{z'_{h,k}}^{z''_{h,k}}\Psi_k(z)e^{\frac{2\pi nz}{k^2}}\,dz \\ &\quad+\sum_{0\leq h<k\leq N} ik^{-\frac{1}{2}}\omega_{h,k}e^{-\frac{2\pi ihn}{k}}\int_{z'_{h,k}}^{z''_{h,k}}\Psi_k(z)\left\{f\left(e^{\frac{2\pi ih'}{k}-\frac{2\pi}{z}}\right)-1\right\}e^{\frac{2\pi nz}{k^2}}\,dz \\ &= \sum_{0\leq h<k\leq N} ik^{-\frac{1}{2}}\omega_{h,k}e^{-\frac{2\pi ihn}{k}}I_{h,k}+\sum_{0\leq h<k\leq N} ik^{-\frac{1}{2}}\omega_{h,k}e^{-\frac{2\pi ihn}{k}}I^*_{h,k}, \end{aligned} \tag{2.6}$$

where $I_{h,k}$ and $I^*_{h,k}$ are respectively abbreviations for the integrals.

3. We estimate first

$$I^*_{h,k} = \int_{z'_{h,k}}^{z''_{h,k}}\Psi_k(z)\left\{f\left(e^{\frac{2\pi ih'}{k}-\frac{2\pi}{z}}\right)-1\right\}e^{\frac{2\pi nz}{k^2}}\,dz. \tag{3.1}$$

The path of integration, which is an arc of the circle K, can here be replaced by the chord $s_{h,k}$ from $z'_{h,k}$ to $z''_{h,k}$. We have on and in the circle K

$$0 < \Re(z) \leq 1, \qquad \Re\left(\frac{1}{z}\right) \geq 1. \tag{3.2}$$

This yields the estimate

$$\left|\Psi_k(z)\left\{f\left(e^{\frac{2\pi ih'}{k}-\frac{2\pi}{z}}\right)-1\right\}e^{\frac{2\pi nz}{k^2}}\right| \leq |z|^{\frac{1}{2}}e^{2\pi n}\sum_{m=1}^{\infty}p(m)e^{-2\pi\left(m-\frac{1}{24}\right)}. \tag{3.3}$$

On the chord $s_{h,k}$ we have $|z|$ less than or equal to the greater of the numbers $|z'_{h,k}|$ and $|z''_{h,k}|$. From (2.3) we derive readily

$$|z'_{h,k}| = \frac{k}{\sqrt{k^2 + k_1^2}}, \qquad |z''_{h,k}| = \frac{k}{\sqrt{k^2 + k_2^2}}.$$

Now

$$\sqrt{k^2 + k_1^2} \geqq 2^{-\frac{1}{2}}(k + k_1) \geqq 2^{-\frac{1}{2}}(N + 1)$$

since k and k_1 are the denominators of neighbor fractions in the Farey series of order N. Therefore we have on $s_{h,k}$

$$|z| \leqq 2^{\frac{1}{2}}k(N + 1)^{-1}. \tag{3.4}$$

The length of the chord $s_{h,k}$ is, according to (2.3),

$$\begin{aligned} k\left\{\left(\frac{k}{k^2 + k_1^2} - \frac{k}{k^2 + k_2^2}\right)^2 + \left(\frac{k_1}{k^2 + k_1^2} - \frac{k_2}{k^2 + k_2^2}\right)^2\right\}^{\frac{1}{2}} \\ = \frac{k\,|k_1 - k_2|}{(k^2 + k_1^2)^{\frac{1}{2}}(k^2 + k_2^2)^{\frac{1}{2}}} \leqq \frac{2k\,|k_1 - k_2|}{(k + k_1)(k + k_2)} \\ = 2k\left|\frac{1}{k + k_1} - \frac{1}{k + k_2}\right| < 2k(N + 1)^{-1}. \end{aligned} \tag{3.5}$$

From (3.1), (3.3), (3.4), (3.5) we obtain now

$$|I^*_{h,k}| < Ck^{\frac{3}{2}}N^{-\frac{3}{2}}$$

where the constant C contains n, which however we keep fixed. This estimate leads to

$$\left|\sum_{0 \leqq h < k \leqq N} ik^{-\frac{1}{2}}\omega_{h,k}e^{-\frac{2\pi ihn}{k}} I^*_{h,k}\right| < CN^{-\frac{3}{2}} \sum_{0 \leqq h < k \leqq N} k^{-1} < CN^{-\frac{1}{2}}.$$

We can thus replace (2.6) by

$$p(n) = \sum_{0 \leqq h < k \leqq N} ik^{-\frac{1}{2}}\omega_{h,k}e^{-\frac{2\pi ihn}{k}} I_{h,k} + O(N^{-\frac{1}{2}}), \tag{3.6}$$

$$I_{h,k} = \int_{z'_{h,k}}^{z''_{h,k}} \Psi_k(z)e^{\frac{2\pi nz}{k^2}}\,dz.$$

4. In $I_{h,k}$ we introduce now the whole circle K, from 0 around in the negative sense to 0, as path of integration:

$$I_{h,k} = \int_{K(-)} \Psi_k(z)e^{\frac{2\pi nz}{k^2}}\,dz - \int_0^{z'_{h,k}} \Psi_k(z)e^{\frac{2\pi nz}{k^2}}\,dz - \int_{z''_{h,k}}^0 \Psi_k(z)e^{\frac{2\pi nz}{k^2}}\,dz.$$

We estimate the last two integrals. Since they are of the same type we need to consider only the first. On the circle K we have

$$\Re\left(\frac{1}{z}\right) = 1, \qquad 0 < \Re(z) \leqq 1. \tag{4.1}$$

The arc from 0 to $z'_{h,k}$ is less than

$$\frac{\pi}{2}\,|z'_{h,k}| < \pi 2^{-\frac{1}{2}} k N^{-1},$$

and (3.4) is also valid on that arc. Remembering the definition (2.5) we obtain therefore

$$\left| \int_0^{z'_{h,k}} \Psi_k(z) e^{\frac{2\pi n z}{k^2}}\, dz \right| < C e^{2\pi n} k^{\frac{3}{2}} N^{-\frac{3}{2}},$$

$$I_{h,k} = \int_{K(-)} \Psi_k(z) e^{\frac{2\pi n z}{k^2}}\, dz + O(k^{\frac{3}{2}} N^{-\frac{3}{2}}).$$

Insertion of this into (3.6) yields

$$\begin{aligned} p(n) &= \sum_{0 \leqq h < k \leqq N} i k^{-\frac{5}{2}} \omega_{h,k} e^{-\frac{2\pi i h n}{k}} \int_{K(-)} \Psi_k(z) e^{\frac{2\pi n z}{k^2}}\, dz \\ &\qquad + O(N^{-\frac{1}{2}}) + \sum_{0 \leqq h < k \leqq N} O(k^{-1} N^{-\frac{3}{2}}) \\ &= i \sum_{1 \leqq k \leqq N} A_k(n) k^{-\frac{5}{2}} \int_{K(-)} \Psi_k(z) e^{\frac{2\pi n z}{k^2}}\, dz + O(N^{-\frac{1}{2}}), \end{aligned}$$

with the abbreviation

$$A_k(n) = \sum_{\substack{h \bmod k \\ (h,k)=1}} \omega_{h,k} e^{-\frac{2\pi i h n}{k}}. \tag{4.2}$$

If we now let N tend to infinity the error term goes to zero, and we obtain an infinite convergent series

$$p(n) = i \sum_{k=1}^{\infty} A_k(n) k^{-\frac{5}{2}} \int_{K(-)} z^{\frac{5}{2}} e^{\frac{\pi}{12 z} + \frac{2\pi z}{k^2}\left(n - \frac{1}{24}\right)}\, dz. \tag{4.3}$$

5. In order to carry out the integral we substitute

$$w = \frac{1}{z}.$$

The path of the integral in the w-plane is then the line parallel to the imaginary axis through the point 1. Therefore we have

$$p(n) = \frac{1}{i} \sum_{k=1}^{\infty} A_k(n) k^{-\frac{5}{2}} \int_{1-i\infty}^{1+i\infty} w^{-\frac{7}{2}} e^{\frac{\pi w}{12} + \frac{2\pi}{k^2}\left(n - \frac{1}{24}\right)\frac{1}{w}}\, dw.$$

The integral here is brought into a known form by the substitution

$$\frac{\pi}{12} w = t,$$

which yields

$$p(n) = \frac{1}{i}\left(\frac{\pi}{12}\right)^{\frac{3}{2}} \sum_{k=1}^{\infty} A_k(n) k^{-\frac{5}{2}} \int_{c-i\infty}^{c+i\infty} t^{-\frac{5}{2}} e^{t+\frac{\pi^2}{6k^2}\left(n-\frac{1}{24}\right)\frac{1}{t}} dt, \qquad c > 0.$$

We find[2] thus

$$p(n) = 2\pi(24n-1)^{-\frac{3}{4}} \sum_{k=1}^{\infty} \frac{A_k(n)}{k} I_{\frac{3}{2}}\left(\frac{\pi}{k}\sqrt{\tfrac{2}{3}\left(n-\tfrac{1}{24}\right)}\right),$$

where $I_{\frac{3}{2}}$ is a "Bessel function of imaginary argument."

Finally, Bessel functions of half odd order can be reduced to elementary functions. In our case we apply the relation

$$I_{\frac{3}{2}}(z) = \sqrt{\frac{2z}{\pi}} \frac{d}{dz}\left(\frac{\sinh z}{z}\right).$$

Using the abbreviations

$$C = \pi\sqrt{\tfrac{2}{3}}, \qquad \lambda_n = \sqrt{n-\tfrac{1}{24}}$$

we obtain then the result (3)

$$p(n) = \frac{1}{\pi\sqrt{2}} \sum_{k=1}^{\infty} A_k(n) k^{\frac{1}{2}} \frac{d}{dn}\left(\frac{\sinh\left(\frac{C}{k}\lambda_n\right)}{\lambda_n}\right),$$

which we had set out to prove.

UNIVERSITY OF PENNSYLVANIA.

REFERENCES

(1) L. R. FORD, *Fractions*. American Mathematical Monthly, vol. 45, (1938), pp. 586–601.

(2) G. H. HARDY AND S. RAMANUJAN, *Asymptotic Formulae in Combinatory Analysis*. Proc. London Math. Soc. (2), vol. 17 (1918), pp. 75–115; also Ramanujan's Collected Papers (1927), pp. 276–309.

(3) H. RADEMACHER, *On the Partition Function* $p(n)$. Proc. London Math. Soc. (2) vol. 43 (1937), pp. 241–254.

---

[2] Watson, Bessel Functions, p. 181, formula (1), where, however, the path of integration is bent into a loop around the negative real axis; compare the remark to formula (8), p. 177.

ANNALS OF MATHEMATICS
Vol. 44, No. 3, July, 1943

SUR LA DISTRIBUTION LIMITE DU TERME MAXIMUM D'UNE SÉRIE ALÉATOIRE

PAR B. GNEDENKO

(Received February 8, 1943)

Introduction

Considérons une suite

$$x_1, x_2, \cdots, x_n, \cdots$$

de variables aléatoires mutuellement indépendantes et assujetties à une même loi de distribution $F(x)$. Formons une autre suite de variables aléatoires

$$\xi_1, \xi_2, \cdots, \xi_n, \cdots,$$

en posant

$$\xi_n = \max(x_1, x_2, \cdots, x_n).$$

On voit facilement que la fonction de distribution de ξ_n est

$$F_n(x) = P\{\xi_n < x\} = F^n(x).$$

L'étude de la fonction $F_n(x)$ pour les grandes valeurs de n offre un intérêt notable. Beaucoup de travaux ont été consacrés à cette question. En particulier, M. Fréchet [1] en 1927 a trouvé les lois qui peuvent être limites pour $F_n(a_n x)$ pour un choix convenable des constantes positives a_n.

Cette classe de lois limites est formée des lois de types[1] suivants

$$\Phi_\alpha(x) = \begin{cases} 0 & \text{pour} \quad x \leqq 0 \\ e^{-x^{-\alpha}} & \text{pour} \quad x > 0 \end{cases}$$

et

$$\Psi_\alpha(x) = \begin{cases} e^{-(-x)^\alpha} & \text{pour} \quad x \leqq 0 \\ 1 & \text{pour} \quad x > 0 \end{cases}$$

où α désigne une constante positive.

En 1928 R. A. Fisher et L. H. C. Tippett [2] ont établi que les lois limites pour $F_n(a_n x + b_n)$, où $a_n > 0$ et b_n sont des constantes réelles convenablement choisies, se réduisent aux lois de types $\Phi_\alpha(x)$, $\Psi_\alpha(x)$ et la loi

$$\Lambda(x) = e^{-e^{-x}}.$$

R. Misès [3] qui a commencé une étude systématique des lois limites pour la distribution du terme maximum vers 1936, a trouvé plusieurs conditions suffisantes pour la convergence des lois $F_n(a_n x + b_n)$ vers chacun des types cités

[1] En imitant A. Khintchine et P. Lévy nous appellerons type de la loi $\Phi(x)$ l'ensemble des lois $\Phi(ax + b)$ pour tous les choix possibles des constantes réelles $a > 0$ et b.

à l'instant, pour un certain choix fixe de constantes $a_n > 0$ et b_n ; en terminant le dernier paragraphe de ce travail nous aurons l'occasion de formuler la condition suffisante de convergence vers la loi $\Lambda(x)$ trouvée par R. Misès.

Cependant les travaux cités ne donnent pas de solution complète des problèmes fondamentaux sur les distributions limites du terme maximum de la série aléatoire. Et, tout d'abord, il reste à rechercher le domaine d'attraction pour chaque loi limite propre[2] $\Phi(x)$, c'est-à-dire l'ensemble de toutes les fonctions de distribution $F(x)$ telles que, pour un choix convenable des constantes $a_n > 0$ et b_n, on ait

$$\lim_{n \to \infty} F^n(a_n x + b_n) = \Phi(x).$$

Il est intéressant de remarquer que non seulement la position des problèmes sur les dirtributions des termes maxima mais, comme nous allons le voir, les résultats obtenus offrent une grande analogie avec les problèmes et les résultats correspondants concernant la théorie des lois stables pour les sommes de variables aléatoires indépendantes (Voir, par exemple, Chap. V [4] et (5)).

Il sera prouvé dans ce travail que la classe des lois limites pour les maxima n'est formée que des types indiqués. Nous donnons aussi les condition nécessaires et suffisantes pour les domaines d'attraction de chacun des types limites possibles. Cependant si les conditions trouvées pour les lois $\Phi_\alpha(x)$ et $\Psi_\alpha(x)$ peuvent être regardées comme définitives, il n'en est pas ainsi pour les conditions concernant la loi $\Lambda(x)$; ces conditions-ci n'ont pas encore reçues, à notre avis, de forme définitive et commode pour les applications. Au §1 sont données les conditions nécessaires et suffisantes pour la loi des grands nombres et pour la stabilité relative des maxima. Il est à remarquez que les lemmes 1 et 2, comme il nous semble, présentent un intérêt par eux-mêmes et peuvent être utiles dans les recherches sur d'autres problèmes limites.

On voit facilement que les resultats que nous allons exposer s'étendent aussi à la distribution du terme minimum de la série aléatoire. Il suffit de remarquer que si

$$\eta_n = \min (x_1, x_2, \cdots, x_n),$$

alors

$$-\eta_n = \max (-x_1, -x_2, \cdots, -x_n).$$

1. La loi des grands nombres

Nous dirons que la suite de maxima

(1) $$\xi_1, \xi_2, \cdots, \xi_n, \cdots$$

[2] Une fonction de distribution s'appelle impropre ou unitaire si elle appartient au type

$$\epsilon(x) = \begin{cases} 0 & \text{pour} \quad x < 0 \\ 1 & \text{pour} \quad x > 0 \end{cases}$$

d'une série de variables aléatoires mutuellement indépendantes

$$x_1, x_2, \cdots, x_n, \cdots \tag{2}$$

est assujettie à la loi des grands nombres s'il existe des constantes A_n telles que l'on ait

$$P\{|\xi_n - A_n| < \epsilon\} \to 1 \tag{3}$$

pour $n \to \infty$ et tout $\epsilon > 0$ donné d'avance.

La suite des maxima (1) sera appellée relativement stable si, pour un choix convenable de constantes positives B_n, la relation

$$\lim_{n\to\infty} P\left\{\left|\frac{\xi_n}{B_n} - 1\right| < \epsilon\right\} = 1 \tag{4}$$

a lieu pour tout $\epsilon > 0$.

Si la fonction de distribution $F(x)$ des variables aléatoires de la suite (2) jouit de cette propiété qu'il existe une valeur x_0 telle que l'on ait

$$F(x_0) = 1 \quad \text{et} \quad F(x_0 - \epsilon) < 1, \tag{5}$$

pour tout $\epsilon > 0$, alors la suite (1) est assujettie à la loi des grands nombres.

En effet, les conditions (5) ayant lieu, nous avons

$$P\{|\xi_n - x_0| < \epsilon\} = 1 - F_n(x_0 - \epsilon) = 1 - F^n(x_0 - \epsilon).$$

Et, puisque pour tout $\epsilon > 0$

$$\lim_{n\to\infty} F^n(x_0 - \epsilon) = 0$$

nous avons pour tout $\epsilon > 0$

$$\lim_{n\to\infty} P\{|\xi_n - x_0| < \epsilon\} = 1.$$

D'une façon analogue, si les conditions (5) ont lieu et si $x_0 > 0$ alors la suite (1) est relativement stable.

En effet, dans ce cas nous avons

$$P\left\{\left|\frac{\xi_n}{x_0} - 1\right| < \epsilon\right\} = 1 - F^n(x_0(1 - \epsilon))$$

et, puisque d'après (5) on a pour $n \to \infty$

$$F^n(x_0(1 - \epsilon)) \to 0,$$

nous obtenons, pour $n \to \infty$,

$$P\left\{\left|\frac{\xi_n}{x_0} - 1\right| < \epsilon\right\} \to 1.$$

Il est évident que dans le cas $x_0 \leqq 0$ la stabilité relative ne peut plus avoir lieu.

Nous voyons ainsi que toute la difficulté concernant la recherche des condi-

tions sous lesquelles ont lieu la loi des grands nombres et la stabilité relative des maxima, se rapporte aux distributions donnant lieu à l'inégalité $F(x) < 1$ pour toutes les valeurs finies de x.

THÉORÈME 1. *Pour que la suite* (1) *soit assujettie à la loi des grands nombres, en supposant* $F(x) < 1$ *pour toute valeur finie de* x, *il faut et il suffit que l'on ait*

$$\lim_{x\to\infty} \frac{1 - F(x + \epsilon)}{1 - F(x)} = 0 \tag{6}$$

pour tout $\epsilon > 0$.

DÉMONSTRATION: En vertu de l'égalité évidente

$$P\{| \xi_n - A_n | < \epsilon\} = F^n(A_n + \epsilon) - F^n(A_n - \epsilon)$$

les conditions pour la loi des grands nombres peuvent être exprimées sous la forme suivante: pour tout $\epsilon > 0$, on a

$$F^n(A_n + \epsilon) \to 1,$$
$$F^n(A_n - \epsilon) \to 0$$

pour $n \to \infty$.

De la première de ces relations il résulte, en tenant compte de la condition du théorème, que $A_n \to \infty$ pour $n \to \infty$.

Les relations trouvées sont équivalentes aux conditions suivantes:

$$n \log F(A_n + \epsilon) \to 0,$$
$$n \log F(A_n - \epsilon) \to -\infty$$

pour $n \to \infty$; or, puisque $1 - F(x) \to 0$ pour $x \to \infty$ et puisque sous cette condition

$$\log F(x) = \log (1 - (1 - F(x))) = -(1 - F(x)) - \tfrac{1}{2}(1 - F(x))^2 - \cdots$$
$$= -(1 - F(x))(1 + 0(1)),$$

les conditions en question sont équivalentes aux relations suivantes

$$\left.\begin{aligned} n(1 - F(A_n + \epsilon)) &\to 0 \\ n(1 - F(A_n - \epsilon)) &\to \infty \end{aligned}\right\} \tag{7}$$

pour $n \to \infty$.

Supposons maintenant les conditions du théorème vérifiées et faisons voir que la loi des grands nombres a lieu. A ce but, définissons les constantes A_n comme les plus petites valeurs de x donnant lieu aux inégalités

$$F(x - 0) \leqq 1 - \frac{1}{n} \leqq F(x + 0). \tag{8}$$

En vertu de l'hypothèse faite sur $F(x)$ dans l'énoncé du théorème il est évident

que $A_n \to \infty$ pour $n \to \infty$. D'après la condition du théorème, nous avons pour tous les ϵ et η ($\epsilon > \eta > 0$)

$$\frac{1 - F(A_n + \epsilon)}{1 - F(A_n + \eta)} \to 0, \qquad \frac{1 - F(A_n + \epsilon)}{1 - F(A_n - \eta)} \to 0,$$

pour $n \to \infty$; or, $\eta > 0$ étant arbitraire, nous en concluons que pour $n \to \infty$

$$(9) \qquad \frac{1 - F(A_n + \epsilon)}{1 - F(A_n + 0)} \to 0, \qquad \frac{1 - F(A_n + \epsilon)}{1 - F(A_n - 0)} \to 0.$$

Il résulte de (8) que

$$\frac{1 - F(A_n + \epsilon)}{1 - F(A_n - 0)} \leqq n(1 - F(A_n + \epsilon)) \leqq \frac{1 - F(A_n + \epsilon)}{1 - F(A_n + 0)},$$

et par conséquent, en vertu de (9), nous avons pour $n \to \infty$

$$(10) \qquad n(1 - F(A_n + \epsilon)) \to 0.$$

De la condition du théorème il résulte que, pour tout $\epsilon > 0$,

$$\lim_{x\to\infty} \frac{1 - F(x)}{1 - F(x - \epsilon)} = 0;$$

nous en tirons par des raisonnements analogues aux précédents que, pour $n \to \infty$, on a

$$(11) \qquad n(1 - F(A_n - \epsilon)) \to \infty.$$

Or, nous avons vu que la relation (3) résulte de (10) et (11).

Supposons maintenant que c'est la loi des grands nombres qui a lieu, c'est-à-dire supposons qu'il existe une suite de constantes A_n telle que, les conditions (10) et (11) soient vérifiées pour tout $\epsilon > 0$. Démontrons alors que l'égalité (6) a lieu elle aussi.

Il est évident que d'après (10) on a $A_n \to \infty$ pour $n \to \infty$, et nous pouvons supposer que les A_n sont non décroissants. Pour tout valeur suffisamment grande de x, nous pouvons trouver un nombre n tel que

$$A_{n-1} \leqq x \leqq A_n.$$

Il est évident que les inégalités

$$1 - F(A_{n-1} - \eta) \geqq 1 - F(x - \eta) \geqq 1 - F(A_n - \eta),$$

$$1 - F(A_{n-1} + \eta) \geqq 1 - F(x + \eta) \geqq 1 - F(A_n + \eta)$$

ont lieu pour tout $\eta > 0$, aussi bien que les inégalités

$$\frac{1 - F(A_{n+1} + \eta)}{1 - F(A_n - \eta)} \geqq \frac{1 - F(x + \eta)}{1 - F(x - \eta)} \geqq \frac{1 - F(A_n + \eta)}{1 - F(A_{n-1} - \eta)}.$$

Il résulte de (10) et (11) que

$$\lim_{x\to\infty} \frac{1 - F(x + \eta)}{1 - F(x - \eta)} = 0.$$

En remplaçant $x - \eta$ par x et 2η par ϵ nous obtenons la condition du théorème.

THÉORÈME 2. *Pour que la suite* (1) *soit relativement stable, en supposant* $F(x) < 1$ *pour toute valeur finie de* x, *il faut et il suffit que la relation*

$$\lim_{x\to\infty} \frac{1 - F(kx)}{1 - F(x)} = 0 \tag{12}$$

ait lieu pour tout $k > 1$.

DÉMONSTRATION: En tenant compte de l'égalité évidente

$$P\left\{\left|\frac{\xi_n}{B_n} - 1\right| < \epsilon\right\} = F^n(B_n(1 + \epsilon)) - F^n(B_n(1 - \epsilon))$$

nous pouvons écrire les conditions de la stabilité sous la forme suivante: pour $n \to \infty$ on a

$$F^n(B_n(1 + \epsilon)) \to 1,$$
$$F^n(B_n(1 - \epsilon)) \to 0.$$

Par des raisonnements analogues à ceux que nous avons employés dans la démonstration du théorème précédent, nous voyons que ces conditions sont équivalentes aux suivantes: pour $n \to \infty$ on a

$$n(1 - F(B_n(1 + \epsilon))) \to 0 \tag{13}$$

$$n(1 - F(B_n(1 - \epsilon))) \to \infty. \tag{14}$$

Supposons d'abord que la condition du théorème soit vérifiée. Définissons B_n comme la plus petite valeur de x donnant lieu aux inégalités

$$F(x(1 - 0)) \leqq 1 - \frac{1}{n} \leqq F(x(1 + 0)). \tag{15}$$

En vertu de l'hypothèse faite sur la fonction $F(x)$ nous concluons que $B_n \to \infty$ pour $n \to \infty$.

De (12) il résulte que pour tous les ϵ et η ($\epsilon > \eta > 0$) on a

$$\frac{1 - F(B_n(1 + \epsilon))}{1 - F(B_n(1 + \eta))} \to 0, \qquad \frac{1 - F(B_n(1 + \epsilon))}{1 - F(B_n(1 - \eta))} \to 0$$

pour $n \to \infty$. Or, puisque $\epsilon > 0$, nous en tirons, pour $n \to \infty$

$$\frac{1 - F(B_n(1 + \epsilon))}{1 - F(B_n(1 + 0))} \to 0, \qquad \frac{1 - F(B_n(1 + \epsilon))}{1 - F(B_n(1 - 0))} \to 0. \tag{16}$$

De l'inégalité (15) nous concluons que

$$\frac{1 - F(B_n(1 + \epsilon))}{1 - F(B_n(1 - 0))} \leqq n(1 - F(B_n(1 + \epsilon))) \leqq \frac{1 - F(B_n(1 + \epsilon))}{1 - F(B_n(1 + 0))},$$

et, par conséquent, en vertu de (16), que

$$n(1 - F(B_n(1 + \epsilon))) \to 0 \tag{17}$$

pour $n \to \infty$. Or, puisqu'il résulte de la condition du théorème que, pour tout $\epsilon > 0$, on a

$$\lim_{x\to\infty} \frac{1 - F(x(1 - \epsilon))}{1 - F(x)} = \infty,$$

nous en obtenons, par des raisonnements analogues aux précédents,

$$n(1 - F(B_n(1 - \epsilon))) \to \infty \tag{18}$$

pour $n \to \infty$.

Mais, comme nous le savons, la stabilité relative des maxima résulte de (17) et (18).

Supposons maintenant que les maxima sont relativement stables et que, par conséquent, les relations (13) et (14) ont lieu. Montrons qu'alors l'égalité (12) a lieu elle-aussi. Du fait que $F(x) < 1$ pour toute valeur finie de x et de la relation (13) il résulte que

$$B_n \to \infty \quad \text{pour} \quad n \to \infty$$

Nous pouvons évidemment supposer que les B_n ne décroissent jamais. Pour toute valeur sufisamment grande de x nous pouvons trouver un entier n tel que

$$B_{n-1} \leqq x \leqq B_n .$$

Il est évident que pour tous les ϵ et $\eta > 0$ nous avons

$$1 - F(B_{n-1}(1 - \eta)) \geqq 1 - F(x(1 - \eta)) \geqq 1 - F(B_n(1 - \eta)),$$
$$1 - F(B_{n-1}(1 + \epsilon)) \geqq 1 - F(x(1 + \epsilon)) \geqq 1 - F(B_n(1 + \epsilon)),$$

d'où l'on tire

$$\frac{1 - F(B_{n-1}(1 + \epsilon))}{1 - F(B_n(1 - \eta))} \geqq \frac{1 - F(x(1 + \epsilon))}{1 - F(x(1 - \eta))} \geqq \frac{1 - F(B_n(1 + \epsilon))}{1 - F(B_{n-1}(1 - \eta))} .$$

Il résulte de (13) et (14) que pour tout $\epsilon > 0$ et tout $\eta > 0$ a lieu l'égalité

$$\lim_{x\to\infty} \frac{1 - F(x(1 + \epsilon))}{1 - F(x(1 - \eta))} = 0.$$

En posant $X = x(1 - \eta)$, $k = \dfrac{1 + \epsilon}{1 - \eta}$, nous obtenons la condition (12).

Considérons à titre d'example les fonctions de distribution suivantes:

$$F_1(x) = \begin{cases} 0 & \text{pour} \quad x \leqq 1 \\ 1 - \dfrac{1}{x^\alpha} & \quad \text{,,} \quad x > 1 \end{cases} \tag{19}$$

$$F_2(x) = \begin{cases} 0 & \text{pour} \quad x \leqq 0 \\ 1 - e^{-x^\alpha} & \quad \text{,,} \quad x > 0 \end{cases} \tag{20}$$

En vertu de

$$\lim_{x\to\infty} \frac{1 - F_1(x+\epsilon)}{1 - F_1(x)} = 1,$$

$$\lim_{x\to\infty} \frac{1 - F_1(kx)}{1 - F_1(x)} = \frac{1}{k^\alpha}$$

nous voyons que pour la fonction de distribution (19) les maximas ne sont pas assujettis a la loi des grands nombres et ne sont pas stables, et cela quel que soit $\alpha > 0$, tandis qu'en vertu de

$$\lim_{x\to 0} \frac{1 - F_2(x+\epsilon)}{1 - F_2(x)} = \begin{cases} 0 & \text{pour} \quad \alpha > 1, \\ e^{-\epsilon} & \text{pour} \quad \alpha = 1, \\ 1 & \text{pour} \quad \alpha < 1, \end{cases}$$

$$\lim_{x\to 0} \frac{1 - F_2(kx)}{1 - F_2(x)} = 0 \qquad (k > 1)$$

nous voyons que pour les lois (20) la stabilité relative a bien lieu pour toutes les valeurs de α, et que la loi des grands nombres n'a lieu que pour $\alpha > 1$.

On vérifie facilement que 1) pour la loi de Poisson les maxima sont relativement stables mais ne sont pas assujettis à la loi des grands nombres, 2) pour la loi de Gauss à dispersion égale a un et à valeur moyenne nulle pour $n \to \infty$ ont lieu les relations suivantes

$$P\left\{\left|\frac{\xi_n}{\sqrt{2 \log n}} - 1\right| < \epsilon\right\} \to 1,$$

$$P\{|\xi_n - \sqrt{2 \log n}| < \epsilon\} \to 1$$

pour tout $\epsilon > 0$.

En 1932 Bruno de Finetti [6] a donné quelques conditions pour l'applicabilité de la loi des grands nombres. Finetti considérait des variables aléatoires ayant densités $f(x) = F'(x)$ et assujetties à certaines conditions supplémentaires; la condition suffisante trouvée par Finetti est exprimée par l'égalité

$$\lim_{z\to\infty} \frac{f(z+\epsilon)}{f(z)} = 0$$

pour tout $\epsilon > 0$. La condition de Finetti résulte facilement du théorème 1 (et cela sans condition supplémentaire imposée aux variables aléatoires). En effet, en admettant l'existence de la dérivée $f(x) = F'(x)$ pour toutes les valeurs de x, on trouve par la règle de l'Hospital

[3] Lorsque j'ai démontré ce théorème, les résultats de Fischer et Tippett exposés dans leur travail cité [2] m'étaient inconnus. Puisque la démonstration de ces auteurs n'est pas, à mon avis, suffisamment détaillée et fait appel à l'hypothèse supperflue de l'analycité des quantités a_n et b_n relativement à la variable n, j'ai pensé qu'il serait utile d'exposer dans ce travail les résultats de ce paragraphe avec tous les développements nécessaires.

$$\lim_{x\to\infty} \frac{1 - F(x + \epsilon)}{1 - F(x)} = \lim_{z\to\infty} \frac{f(z + \epsilon)}{f(z)}$$

d'où il résulte, en vertu du théoreme 1 de ce travail, que dans le cas de l'existence de la limite

$$\lim \frac{f(z + \epsilon)}{f(z)}$$

pour tout $\epsilon > 0$ la condition de Finetti est nécessaire et suffisante.

Il est évident qu'on peut énoncer une condition analogue concernant la stabilité relative des maxima.

2. La classe des lois limites

THÉORÈME 3. *La classe des lois limites pour $F_n(a_n x + b_n)$ où $a_n > 0$ et b_n sont des constantes convenablement choisies, ne contient que les lois des types $\Phi_\alpha(x)$, $\Psi_\alpha(x)$, $\Lambda(x)$.*[3]

DÉMONSTRATION: Supposons que pour un choix de constantes $a_n > 0$ et b_n on ait

$$F_n(a_n x + b_n) = F^n(a_n x + b_n) \to \Phi(x)$$

pour $n \to \infty$. Alors, l'égalité

$$\lim_{n\to\infty} [F^n(a_{nk} x + b_{nk})]^k = \Phi(x) \tag{21}$$

a lieu pour entier $k > 0$. Il en résulte que, k étant constant et $n \to \infty$, la suite des fonctions $F^n(a_{nk}x + b_{nk})$ tend vers une fonction limite. En vertu d'un théorème de A. Khintchine ([4], théorème 43) cette fonction limite doit appartenir au même type que la fonction $\Phi(x)$, c'est-à-dire que nous devons avoir

$$\lim_{n\to\infty} F^n(a_{nk} x + b_{nk}) = \Phi(\alpha_k x + \beta_k), \tag{22}$$

où les $\alpha_k > 0$ et β_k sont constantes.

Il résulte de (21) et de (22) que pour tout nombre naturel k la loi limite satisfait à l'égalité

$$\Phi^k(\alpha_k x + \beta_k) = \Phi(x). \tag{23}$$

Considérons séparement les trois cas suivants.

1) $\alpha_k < 1$ pour une certaine valeur de $k > 1$. Pour

$$x \geqq \frac{\beta_k}{1 - \alpha_k}$$

nous avons

$$\alpha_k x + \beta_k \leqq x.$$

Donc, la fonction $\Phi(x)$ étant monotone, nous pouvons écrire

$$\Phi(\alpha_k x + \beta_k) \leqq \Phi(x).$$

Il en résulte que pour la fonction de distribution $\Phi(x)$ l'égalité (23) ne peut avoir lieu que si

$$\Phi(x) = 1 \qquad \text{pour} \qquad x \geqq \frac{\beta_k}{1 - \alpha_k}.$$

Montrons maintenant que pour $x < \beta_k/(1 - \alpha_k)$ on doit avoir

$$\Phi(x) < 1.$$

Supposons le contraire, c'est-à-dire admettons qu'il existe une valeur $x_0 < \beta_k/(1 - \alpha_k)$ donnant lieu a l'inégalité

$$\Phi(x_0) = 1. \tag{24}$$

Il est évident qu'il est toujours possible de choisir pour tout $x < x_0$ un entier n tel que

$$x_0 \leqq \alpha_k^n x + \beta_k(1 + \alpha_k + \cdots + \alpha_k^{n-1}).$$

Alors, en vertu de (24), on doit avoir

$$\Phi(\alpha_k^n x + \beta_k(1 + \alpha_k + \cdots + \alpha_k^{n-1})) = 1.$$

Or, il résulte de (23) que

$$\begin{aligned} \Phi^{k^n}(\alpha_k^n + \beta_k(1 + \alpha_k + \cdots + \alpha_k^{n-1})) \\ &= [\Phi^k(\alpha_k\{\alpha_k^{n-1} + \beta_k(1 + \alpha_k + \cdots + \alpha_k^{n-2})\} + \beta_k)]^{k^{n-1}} \\ &= \Phi^{k^{n-1}}(\alpha_k^{n-1} + \beta_k(1 + \alpha_k + \cdots + \alpha_k^{n-2})) = \Phi(x) \end{aligned} \tag{25}$$

c'est-à-dire que

$$\Phi(x) = 1$$

pour toute valeur de x, ce qui est impossible.

Nour voyons ainsi que $\Phi(x) \to 1$ pour $x \geqq \beta_k/(1 - \alpha_k)$ et $\Phi(x) < 1$ pour $x < \beta_k/(1 - \alpha_k)$.

Montrons maintenant que si $\Phi(x)$ est une loi propre et si $\alpha_k < 1$ pour une valeur de k, cette égalité a lieu aussi pour toutes les valeurs de k. Admettons qu'il existe un nombre $r > 1$ donnant lieu à l'inégalité $a_r \geqq 1$.

Si $\alpha_r = 1$, nous avons pour toutes les valeurs de x, $\Phi^r(x + \beta_r) = \Phi(x)$ et, par conséquent, $\Phi^r(x) = \Phi(x - \beta_r)$. D'où il résulte en particulier que

$$\left\{ \begin{aligned} \Phi^r\left(\frac{\beta_k}{1 - \alpha_k} + \beta_r\right) = \Phi\left(\frac{\beta_k}{1 - \alpha_k}\right) = 1, \\ \Phi\left(\frac{\beta_k}{1 - \alpha_k} - \beta_r\right) = \Phi^r\left(\frac{\beta_k}{1 - \alpha_k}\right) = 1. \end{aligned} \right. \tag{26}$$

Si $\beta_r \neq 0$, nous avons min $(\beta_k/(1 - \alpha_k) + \beta_r; \beta_k/(1 - \alpha_k) - \beta_r) < \beta_k/(1 - \alpha_k)$ et il résulte de (26) en vertu des raisonnements développés à l'instant que $\Phi(x) \equiv 1$; si c'est le cas $\beta_r = 0$ qui se présente, alors nous avons

$$\Phi^2(x) = \Phi(x)$$

et $\Phi(x) \equiv 1$ ou $\Phi(x) = 1$ pour $x \geqq \beta_k/(1 - \alpha_k)$, tandis que $\Phi(x) = 0$ pour $x < \beta_k/(1 - \alpha_k)$. Donc, $\Phi(x)$ étant une loi propre et α_k vérifiant l'inégalité $\alpha_k < 1$, nous voyons que $\alpha_r \neq 1$ pour tous les r.

Si $\alpha_r > 1$, on a pour $x \leqq \beta_r/(1 - \alpha_r)$

$$\alpha_r x + \beta_r \leqq x$$

et, par conséquent,

$$\Phi(\alpha_r x + \beta_r) \leqq \Phi(x).$$

D'où, en vertu de (23), nous tirons

$$\Phi(x) = 0 \qquad \text{pour} \qquad x \leqq \frac{\beta_r}{1 - \alpha_r}. \tag{27}$$

Soit $x < \beta_k/(1 - \alpha_k)$. Pour tout $\epsilon > 0$ on peut trouver un n tel que

$$\frac{\beta_k}{1 - \alpha_k} - \epsilon < \alpha_k^n x + \beta_k(1 + \alpha_k + \cdots + \alpha_k^{n-1}) = z.$$

Conformément à (25) et (27) nous obtenons

$$\Phi^{k^n}(\alpha_k^n x + \beta_k(1 + \alpha_k + \cdots + \alpha_k^{n-1})) = \Phi(x) = 0,$$

c'est-à-dire que pour tout $z < \beta_k/(1 - \alpha_k)$ nous avons

$$\Phi(z) = 0.$$

La loi $\Phi(x)$ est donc impropre, ce qui est en contradiction avec l'hypothèse admise.

La fonction $\Phi(x)$ est donc telle que $\Phi(x) = 1$ pour $x \geqq \beta_k/(1 - \alpha_k) = x_0$ et $\Phi(x) < 1$ ($\neq 0$) pour $x < \beta_k/(1 - \alpha_k) = x_0$. Or, le point x_0 étant évidemment indépendant de k, nous avons

$$\frac{\beta_k}{1 - \alpha_k} = \frac{\beta_n}{1 - \alpha_n}$$

pour toutes les valeurs de k et de n.

Posons

$$\bar{\Phi}(z) = \Phi\left(z + \frac{\beta_k}{1 - \alpha_k}\right)$$

(cela correspond au déplacement de l'origine au point $\beta_k/(1 - \alpha_k)$).

Il est évident que

$$\bar{\Phi}(\alpha_k z) = \Phi\left(\alpha_k z + \frac{\beta_k}{1 - \alpha_k}\right).$$

En vertu de (23) la fonction $\bar{\Phi}(z)$ vérifie l'équation

$$\bar{\Phi}^k(\alpha_k z) = \bar{\Phi}(z) \tag{28}$$

pour tout entier positif k. La solution de cette équation fonctionnelle est bien connue (voir, par exemple, [4], page 95) la seule fonction de distribution vérifiant l'équation (28) et assujettie à la condition $\bar{\Phi}(z) = 1$ pour $z \geqq 0$ est la fonction $\Psi_\alpha(x)$.

2) $\alpha_k > 1$ pour un certain k. Il résulte des raisonnements exposés que $\alpha_k > 1$ pour toutes les valeurs de k. Nous avons déjà vu (27) que

$$\Phi(x) = 0 \qquad \text{pour} \qquad x \leqq \frac{\beta_k}{1 - \alpha_k}.$$

La démonstration de l'inégalité $\Phi(x) > 0$ pour $x > \beta_k/(1 - \alpha_k)$ résule de l'égalité

$$\Phi^{k^n}(x) = \Phi\left(\frac{x}{\alpha_k^n} - \beta_k\left(\frac{1}{\alpha_k} + \frac{1}{\alpha_k^2} + \cdots + \frac{1}{\alpha_k^n}\right)\right)$$

qui résulte facilement à son tour de (23). De cette même inégalité il résulte que $\Phi(x) < 1$ pour tous les $x > \beta_k/(1 - \alpha_k)$. D'une façon semblable nous voyons que

$$\frac{\beta_k}{1 - \alpha_k} = \frac{\beta_n}{1 - \alpha_n}$$

pour toutes les valeurs de k et n, et que la fonction

$$\bar{\Phi}(z) = \Phi\left(z + \frac{\beta_k}{1 - \alpha_k}\right)$$

vérifie, pour tous les $k > 0$, l'équation (28).

La seule fonction de distribution solution de cette équation et satisfaisant à la condition $\bar{\Phi}(z) = 0$ pour $z < 0$ est la fonction $\Phi_\alpha(x)$.

3) $\alpha_k = 1$ pour un certain k. Il résulte de ce qui précède que $\alpha_k = 1$ pour tous les k. En effectuant le changement de variable

$$z = e^x, \qquad \beta_k = e^{c_k}, \qquad \bar{\Phi}(z) = \begin{cases} \Phi(\log z) & \text{pour} \quad z > 0 \\ 0 & \text{pour} \quad z \leqq 0 \end{cases}$$

nous réduisons l'équation (23) à la forme

$$\bar{\Phi}^k(c_k z) = \bar{\Phi}(x)..$$

La seule fonction de distribution satisfaisant a cette équation et à la condition $\bar{\Phi}(0) = 0$ est la fonction $\Phi_\alpha(z)$. Ainsi, nous avons

$$\Phi(x) = e^{-e^{-\alpha x}}.$$

Cette fonction est du type $\Lambda(x)$, dont nous avons parlé dans l'Introducion.

3. Propositions auxiliaires

LEMME 1. *Soient $F_n(x)$ et $\Phi(x)$ des fonctions de distribution, $\Phi(x)$ n'étant pas unitaire. Si pour certaines suites de nombres réels $a_n > 0$, b_n, $\alpha_n > 0$, β_n on a*

$$F_n(a_n x + b_n) \to \Phi(x)$$

et

$$F_n(\alpha_n x + \beta_n) \to \Phi(x)$$

pour $n \to \infty$ on aura pour $n \to \infty$

$$\frac{a_n}{\alpha_n} \to 1, \qquad \frac{b_n - \beta_n}{a_n} \to 0.$$

DÉMONSTRATION: Posons pour abréger

$$V_n(x) = F_n(a_n x + b_n)$$

D'après l'hypothèse du lemme nous avons pour $n \to \infty$

$$V_n(x) \to \Phi(x)$$

et

$$V_n(A_n x + B_n) \to \Phi(x),$$

où

$$A_n = \frac{\alpha_n}{a_n}, \qquad B_n = \frac{\beta_n - b_n}{a_n}.$$

Déterminons une suite d'indices $n_1 < n_2 < \cdots < n_k < \cdots$ telle que les limites

$$\lim_{k\to\infty} A_{n_k} = A, \qquad \lim_{k\to\infty} B_{n_k} = B \qquad (0 \leqq A \leqq +\infty, -\infty \leqq B \leqq +\infty)$$

existent. Montrons que $A < +\infty$. Supposons le contraire, c'est à-dire que $A = +\infty$, et désignons par x_0 la borne supérieure des nombres x pour lesquels

$$\overline{\lim_{k\to\infty}} (A_{n_k} x + B_{n_k}) < +\infty.$$

Il est évident que pour tout $x > x_0$ on a

$$\overline{\lim_{k\to\infty}} (A_{n_k} x + B_{n_k}) = +\infty,$$

tandis que pout tout $x < x_0$

$$\overline{\lim_{k\to\infty}} (A_{n_k} x + B_{n_k}) = -\infty.$$

Nous en tirons que

$$\Phi(x) = \begin{cases} 0 & \text{pour} \quad x < x_0 \\ 1 & \text{pour} \quad x > x_0. \end{cases}$$

Mais ceci étant exclu par l'hypothèse du lemme, nous avons $A < +\infty$. Il en résulte que B est fini lui aussi, puisque de l'hypotèse $B = -\infty$ résulterait pour toutes les valeurs de x l'égalité

$$\overline{\lim_{k\to\infty}} (A_{n_k} x + B_{n_k}) = -\infty,$$

d'ou il résulterait, à son tour, que $\Phi(x) \equiv 0$; de même l'hypothèse $B = +\infty$ entrainerait

$$\underline{\lim}_{k\to\infty} (A_{n_k} x + B_{n_k}) = +\infty,$$

d'où il résulterait que $\Phi(x) \equiv 1$. Il est évident que $A > 0$, puisque les α_n et les a_n jouent le même rôle, et s'il était $A = 0$, nous aurions la relation $\overline{\lim}_{k\to\infty} (a_{n_k}/\alpha_{n_k}) = +\infty$, dont l'impossibilité vient d'être démontrée.

Soit x une valeur telle qu' aux points x et $Ax + B$ la fonction $\Phi(x)$ soit continue. Il est alors évident que

$$\Phi(x) = \lim_{k\to\infty} V_{n_k}(A_{n_k} x + B_{n_k}) = \Phi(Ax + B). \tag{29}$$

Cette égalité devant avoir lieu pour tous les x, sauf les points de discontinuité au plus, il en résulte que $A = 1$ et $B = 0$. En effet, admettons qu'il n'en n'est pas ainsi et considérons les cas qui peuvent alors se présenter.

Pour $A < 1$, en itérant (29), nous obtenons pour tous les x et pour n naturel et arbitraire

$$\Phi(x) = \Phi(A^n x + B(1 + A + \cdots + A^{n-1})).$$

Et puisque $A^n x$, pour n suffisamment grand, peut être rendu aussi petit que l'on veut, nous avons pour tous les x

$$\Phi(x) = \lim_{n\to\infty} \Phi(A^n x + B(1 + A + \cdots + A^{n-1})) = \Phi\left(\frac{B}{1 - A}\right).$$

La fonction $\Phi(x)$ n'est donc pas une fonction de distribution.

Pour $A > 1$, nous écrivons (29) dans la forme

$$\Phi(x) = \Phi\left(\frac{x}{A} - \frac{B}{A}\right),$$

et par des raisonnements analogues aux précédents nous arrivons à l'égalité

$$\Phi(x) = \Phi\left(\frac{B}{1 - A}\right)$$

qui doit avoir lieu pour toutes les valeurs de x, ce qui prouve que $\Phi(x)$ ne peut pas être une fonction de distribution. Prouvons enfin l'impossibilité du cas $A = 1$, $B \neq 0$. En effet, en vertu de (29) nous voyons que pour tout entier n on a

$$\Phi(x) = \Phi(x + nB).$$

Si $B > 0 (B < 0)$ alors, pour tous les x et pour $n \to \infty (n \to -\infty)$

$$\Phi(x) = \Phi(x + nB) \to \Phi(+\infty)$$

d'autre part, pour $n \to -\infty \ (n \to +\infty)$

$$\Phi(x) = \Phi(x + nB) \to \Phi(-\infty),$$

ce qui n'est évidemment possible que si $\Phi(x)$ est constante, et, par conséquent, n'est pas une fonction de distribution. On voit donc que lorsque pour une suite d'indices $\{n_k\}$ les limites

$$\lim_{n\to\infty} A_{n_k} = A, \qquad \lim_{k\to\infty} B_{n_k} = B$$

existent, on a nécessairement $A = 1$, $B = 0$. Or, nous avons à démontrer que

$$\lim_{n\to\infty} A_n = 1, \qquad \lim_{n\to\infty} B_n = 0$$

si l'une ou l'autre de ces relations n'avait par lieu, on aurait évidemment une suite d'indices $\{n_k\}$ et un nombre $\epsilon > 0$ tels qu'une au moins des inéglalités

$$\varliminf_{k\to\infty} | A_{n_k} - 1 | \geqq \epsilon, \qquad \varliminf_{k\to\infty} | B_{n_k} | \geqq \epsilon \tag{30}$$

serait remplie; de plus, la suite des n_k peut être choisie de manière que A_{n_k} et B_{n_k} tendent vers des limites fixes pour $k \to \infty$; or, ces limites ne peuvent être que 1 et 0 respectivement, d'après ce que nous venons de démontrer; ce résultat étant en contradiction avec (30), le lemme est démontré.

Nous aurons aussi à faire usage dans la suite de la proposition inverse.

LEMME 2. *Si $F_n(x)$ est une suite de fonctions de distribution donnant lieu à la relation*

$$\lim_{n\to\infty} F_n(a_n x + b_n) = \Phi(x):$$

pour un certain choix de constantes $a_n > 0$ et b_n et pour toutes les valeurs de x, alors pour deux suites quelconques de constantes $\alpha_n > 0$ et β_n telles que pour $n \to \infty$

$$\frac{a_n}{\alpha_n} \to 1, \qquad \frac{b_n - \beta_n}{\alpha_n} \to 0 \tag{31}$$

on a

$$F_n(\alpha_n x + \beta_n) \to \Phi(x)$$

pour $n \to \infty$ et toutes les valeurs de x.

DÉMONSTRATION: Soient x_1, x et x_2 $(x_1 < x < x_2)$ des points de continuité de la fonction $\Phi(x)$. En vertu de notre hypothèse nous avons pour n assez grand

$$x_1 < \frac{\alpha_n}{a_n} x + \frac{\beta_n - b_n}{a_n} < x_2 .$$

Puisque

$$\alpha_n x + \beta_n = a_n\left(\frac{\alpha_n}{a_n} x + \frac{\beta_n - b_n}{a_n}\right) + b_n ,$$

on a, pour n assez grand,

$$a_n x_1 + b_n < \alpha_n x + \beta_n < a_n x_2 + b_n$$

et par suite

$$F_n(a_n x_1 + b_n) \leqq F_n(a_n x + b_n) \leqq F_n(a_n x_2 + b_n).$$

Tenant compte de (31) ceci montre que

$$\Phi(x_1) \leqq \varliminf_{n\to\infty} F_n(\alpha_n x + \beta_n) \leqq \varlimsup_{n\to\infty} F_n(\alpha_n x + \beta_n) \leqq \Phi(x_2).$$

En faisant x_1 et x_2 tandre vers x, on aura

$$\Phi(x_1) \to \Phi(x), \qquad \Phi(x_2) \to \Phi(x)$$

x étant supposé point de continuité de $\Phi(x)$. On a donc

$$\Phi(x) \leqq \varliminf_{n\to\infty} F_n(\alpha_n x + \beta_n) \leqq \varlimsup_{n\to\infty} F_n(\alpha_n x + \beta_n) \leqq \Phi(x),$$

c.q.f.d.

LEMME 3. *Si $F(x)$ est une fonction de distribution et si pour un choix de constantes $a_n > 0$ et b_n on a, pour $n \to \infty$ et toutes les valeurs de x*

(32) $$F^n(a_n x + b_n) \to \Phi(x),$$

$\Phi(x)$ etant une fonction de distribution propre, alors on a pour $n \to \infty$,

$$\frac{a_n}{a_{n+1}} \to 1, \qquad \frac{b_n - b_{n+1}}{a_n} \to 0.$$

DÉMONSTRATION. En effet, si la relation (32) a lieu, nous avons pour toutes les valeurs de x telles que $\Phi(x) \neq 0$

$$\lim_{n\to\infty} F(a_n x + b_n) = 1.$$

D'où l'on tire que, pour $n \to \infty$

$$F^{n+1}(a_n x + b_n) \to \Phi(x)$$

Nous nous trouvons donc dans les conditions du lemme 1 avec $\alpha_n = a_{n-1}$, $\beta_n = b_{n-1}$, et ceci démontre le lemme en question.

LEMME 4. *Pour que l'on ait*

(33) $$F^n(a_n x + b_n) \to \Phi(x)$$

pour toutes les valeurs de x et pour $n \to \infty$, il faut et il suffit d'avoir pour $n \to \infty$

(34) $$n[1 - F(a_n x + b_n)] \to -\log \Phi(x)$$

pour toutes les valeurs de x telles que $\Phi(x) \neq 0$.

DÉMONSTRATION: Supposons que la relation (33) ait lieu; alors pour toute valeur de a telle que $\Phi(x) \neq 0$, nous avons

$$\lim_{n\to\infty} F(a_n x + b_n) = 1 \tag{35}$$

Il est évident que pour ces valeurs de x la condition (33) équivaut à la suivante

$$n \log F(a_n x + b_n) \to \log \Phi(x) \qquad (\Phi(x) \neq 0) \tag{36}$$

pour $n \to \infty$. Or, en vertu de (35), nous avons

$$\begin{aligned} \log F(a_n x + b_n) &= -(1 - F(a_n x + b_n)) - \tfrac{1}{2}(1 - F(a_n x + b_n))^2 \\ &- \cdots = -(1 - F(a_n x + b_n)) \quad (1 + 0(1)). \end{aligned} \tag{37}$$

D'où nous voyons que, dès que (33) a lieu, la condition (34) est nécessairement remplie. Inversement, si c'est la condition (34) qui a lieu, alors (35) a lieu aussi, donc, en vertu de (37),

$$-n[1 - F(a_n x + b_n)] = n \log F(a_n x + b_n) \qquad (1 + 0(1)).$$

D'où, et en vertu de (34), résulte (36) et, par conséquent, (33).

4. Le domaine d'attraction de la loi $\Phi_\alpha(x)$

THÉORÈME 4. *Pour qu'une fonction de distribution $F(x)$ appartienne au domaine d'attration de la loi $\Phi_\alpha(x)$ il faut et ill suffit qu l'on ait*

$$\lim_{x\to\infty} \frac{1 - F(x)}{1 - F(kx)} = k^\alpha \tag{38}$$

pour toute valuer de $k > 0$.

DÉMONSTRATION: Supposons d'abord que la condition (38) est vérifiée et montrons que la fontion $F(x)$ appartient au domaine d'attraction de la loi $\Phi_\alpha(x)$.

Il est évident, d'après (38), que $F(x) < 1$ pour toute valeur de x. Il en résulte que, pour n suffisamment grand, les valeurs de x donnant lieu à l'inégalité

$$1 - F(x) \leqq \frac{1}{n}$$

sont positives.

Définissons a_n comme la plus petite des valeurs de x vérifiant les inégalités

$$1 - F(x(1 + 0)) \leqq \frac{1}{n} \leqq 1 - F(x(1 - 0)) \tag{39}$$

il résulte de ce qui précede que $a_n \to \infty$ pour $n \to \infty$. D'après la condition du théorème, pour toute valeur de x et pour tout $\epsilon(0 < \epsilon < 1)$ nous avons

$$\frac{1 - F(a_n x)}{1 - F(a_n(1 + \epsilon))} \to \left(\frac{1 + \epsilon}{x}\right)^\alpha$$

et

$$\frac{1 - F(a_n x)}{1 - F(a_n(1 - \epsilon))} \to \left(\frac{1 - \epsilon}{x}\right)^{\alpha}$$

n tendant vers l'infini. Les premiers membres de ces relations étant fonctions monotones de ϵ et les seconds membres étant fonctions continues de ϵ, la convergence en question est uniforme, ce qui nous permet d'écrire, pour $n \to \infty$,

$$\frac{1 - F(a_n x)}{1 - F(a_n(1 + 0))} \to \frac{1}{x^{\alpha}}$$

et

$$\frac{1 - F(a_n x)}{1 - F(a_n(1 - 0))} \to \frac{1}{x^{\alpha}}.$$

Or, puisque d'après (39) nous avons

$$\frac{1 - F(a_n x)}{1 - F(a_n(1 - 0))} \leqq n(1 - F(a_n x)) \leqq \frac{1 - F(a_n x)}{1 - F(a_n(1 + 0))},$$

on voit que pour tout $x > 0$ on a

$$n(1 - F(a_n x)) \to x^{-\alpha}$$

pour $n \to \infty$.

D'après le Lemme 4 du paragraphe précédent on a pour $n \to \infty$

$$F^n(a_n x) \to \Phi_\alpha(x) \qquad (-\infty < x < +\infty).$$

Supposons maintenant que $F(x)$ appartient au domaine d'attraction de la loi $\Phi_\alpha(x)$, c'est-à-dire supposons que, pour un choix convenable de constantes $a_n > 0$ et b_n, pour toutes les valeurs de $x (x > 0)$ a lieu la relation

(40) $$n(1 - F(a_n x + b_n)) \to x^{-\alpha}$$

pour $n \to \infty$, et faisons voir que la condition (38) en résulte.

Pour toute constante $\beta > 1$ nous avons pour $n \to \infty$

$$n\beta(1 - F(a_n x + b_n)) \to \beta x^{-\alpha}.$$

Puisqu'il résulte de (40) que pour $x > 0$ et $n \to \infty$ on a

$$1 - F(a_n x + b_n) \to 0$$

nous voyons que, pour $n \to \infty$, on doit avoir

$$[n\beta](1 - F(a_n x + b_n)) \to \beta x^{-\alpha},$$

où $[n\beta]$ désigne l'entier de $n\beta$.

Par le changement de variable $x = z\beta^{1/\alpha}$ cette relation prend la forme suivante: pour $n \to \infty$ et pour tout $z > 0$ on a

(41) $$[n\beta](1 - F(a_n \beta^{1/\alpha} z + b_n)) \to z^{-\alpha}.$$

Il résulte de (3) que pour $n \to \infty$

(42) $$[n\beta](1 - F(a_{[n\beta]}x + b_{[n\beta]})) \to x^{-\alpha}$$

En vertu des Lemmes 4 et 1 et en tenant compte de (4) et (5), nous concluons que les relations

$$\frac{a_{[n\beta]}}{a_n} \to \beta^{1/\alpha}, \qquad \frac{b_{[n\beta]} - b_n}{a_{[n\beta]}} \to 0$$

doivent avoir lieu pour $n \to \infty$. En vertu du Lemme 2 la relation (42) ne changera pas si nous posons

(43) $$a_{[n\beta]} = a_n\beta^{1/\alpha}, \qquad b_{[n\beta]} = b_n .$$

Posons

$$n_s = [n_{s-1,\beta}], \qquad n_1 = [n\beta]$$

où l'entier n est à considérer comme fixe. Il résulte de (43) que pour tout nombre naturel s nous avons

$$a_{n_s} = a_n\beta^{s/\alpha}, \qquad b_{n_s} = b_n .$$

D'où nous tirons que pour $s \to \infty$

$$\frac{b_{n_s}}{a_{n_s}} \to 0$$

et par conséquent, en vertu du Lemme 2, que pour $s \to \infty$

(44) $$n_s[1 - F(a_{n_s}x)] \to x^{-\alpha}$$

Supposons que $y \to \infty$; pour toute valeur suffisamment grande de y ou peut trouver une valeur de s telle que

$$a_{n_s}x \leqq y \leqq a_{n_{s+1}}x$$

et, par suite

$$1 - F(a_{n_{s+1}}x) \leqq 1 - F(y) \leqq 1 - F(a_{n_s}x)$$

et, pour $k > 0$

$$1 - F(a_{n_{s+1}}kx) \leqq 1 - F(ky) \leqq 1 - F(a_{n_s}kx).$$

Nous en tirons l'inégalité

(45) $$\frac{1 - F(a_{n_{s+1}}x)}{1 - F(a_{n_s}kx)} \leqq \frac{1 - F(y)}{1 - F(ky)} \leqq \frac{1 - F(a_{n_s}x)}{1 - F(a_{n_{s+1}}kx)}$$

Remarquons que

$$\frac{n_{s+1}}{n_s} = \frac{n_s\beta - \theta_s}{n_s},$$

où $0 \leqq \theta_s < 1$; donc pour $s \to \infty$ on a

$$\frac{n_{s+1}}{n_s} \to \beta.$$

Nous en concluons, en vertu de (44) et (45), que

$$\frac{1}{\beta} k^\alpha \leqq \lim_{y \to \infty} \frac{1 - F(y)}{1 - F(ky)} \leqq \beta k^\alpha;$$

or, puisque β peut être choisi aussi peu différent de l'unité que l'on veut, la condition du théorème en résulte.

Faisons remarquer qu'il résulte de ce qui précède que toute fonction de distribution $F(x)$ appartenant au domaine d'attraction de la loi $\Phi_\alpha(x)$ est attirée vers $\Phi_\alpha(x)$ d'une façon plus particulière, à savoir: pour un choix de constantes a_n a lieu l'égalité

$$\lim_{n \to \infty} F^n(a_n x) = \Phi_\alpha(x).$$

5. Le domaine d'attraction de la loi $\Psi_\alpha(x)$

THÉORÈME 5. *Pour qu'une fonction de distribution $F(x)$ appartienne au domaine d'attraction de la loi $\Psi_\alpha(x)$ il faut et il suffit que*

1. *il existe un x_0 tel que*

$$F(x_0) = 1 \quad \text{et} \quad F(x_0 - \epsilon) < 1$$

pour tout $\epsilon > 0$.

2. $$\lim_{x \to -0} \frac{1 - F(kx + x_0)}{1 - F(x + x_0)} = k^\alpha$$

pour tout $k > 0$.

DÉMONSTRATION: Supposons que les conditions du théorème ont lieu et montrons que la fonction $F(x)$ appartient au domaine d'attraction de la loi $\Psi_\alpha(x)$. A cette fin nous définissons a_n comme la plus petite des valeurs $x > 0$ donnant lieu aux inégalités

$$1 - F(-x(1 - 0) + x_0) \leqq n \leqq 1 - F(-x(1 + 0) + x_0). \tag{46}$$

D'après la première condition du théorème nous avons pour $n \to \infty$

$$a_n \to 0.$$

La deuxième condition du théorème nous fournit les relations

$$\frac{1 - F(a_n x + x_0)}{1 - F(-a_n(1 + \epsilon) + x_0)} \to \left(-\frac{x}{1 + \epsilon}\right)^\alpha$$

et

$$\frac{1 - F(a_n x + x_0)}{1 - F(-a_n(1 - \epsilon) + x_0)} \to \left(-\frac{x}{1 - \epsilon}\right)^\alpha$$

pour tout $\epsilon > 0$ et $x < 0$, n tendant vers l'infini. Or, les premiers membres de ces relations sont des fonctions monotones de ϵ et les seconds membres des fonctions continues de ϵ; donc la convergence dont il s'agit est uniforme, ce qui nous permet d'écrire pour $n \to \infty$

$$\frac{1 - F(a_n x + x_0)}{1 - F(-a_n(1 + 0) + x_0)} \to (-x)^\alpha$$

et

$$\frac{1 - F(a_n x + x_0)}{1 - F(-a_n(1 - 0) + x_0)} \to (-x)^\alpha.$$

Or, puisqu'en vertu de (1) nous avons

$$\frac{1 - F(a_n x + x_0)}{1 - F(-a_n(1 + 0) + x_0)} \leqq n(1 - F(a_n x + x_0)) \leqq \frac{1 - F(a_n x + x_0)}{1 - F(-a_n(1 - 0) + x_0)},$$

nous pouvons affirmer que pour tous les $x < 0$ et pour $n \to \infty$ on a aussi

$$n(1 - F(a_n x + x_0)) \to (-x)^\alpha$$

d'où, en vertu du Lemme 4, on voit que la fonction $F(x)$ appartient au domaine d'attraction de la loi $\Psi_\alpha(x)$.

Supposons maintenant que $F(x)$ appartient au domaine d'attraction de la loi $\Psi_\alpha(x)$, ce qui veut dire que pour toutes les valeurs de x et pour un certain choix des $a_n > 0$ et b_n, on a

$$F^n(a_n x + b_n) \to \Psi_\alpha(x), \tag{47}$$

lorsque $n \to \infty$. Nous en tirons, pour $n \to \infty$

$$F^{2n}(a_n x + b_n) \to \Psi_\alpha^2(x) = \Psi_\alpha(\gamma x), \tag{48}$$

où

$$\gamma = 2^{1/\alpha}.$$

En remplaçant dans (48) γx par x, nous voyons que, pour $n \to \infty$ on a

$$F^{2n}\left(\frac{a_n}{\gamma} x + b_n\right) \to \Psi_\alpha(x). \tag{49}$$

En comparant (47) et (49), nous concluons en vertu des Lemmes 1 et 2 que les a_n et b_n peuvent bien être choisies de manière à avoir

$$a_{2n} = \frac{a_n}{\gamma}, \qquad b_{2n} = b_n.$$

D'où il résulte que nous pouvons toujours faire ce choix de façon à avoir, pour tout k,

$$b_{2^k n} = b_{2^{k-1} n} = b_n \tag{50}$$

et que $a_n \to 0$ pour $n \to \infty$.

Si l'inégalité

$$F(x) < 1$$

a lieu pour toutes les valeurs de x, nous tirons de (47), en y posant $x = 0$, que $b_n \to \infty$ pour $n \to \infty$, ce qui est en contradiction avec (50). Nous avons donc démontré la nécessité de la première condition du théorème. Si la relation (47) a lieu, on doit choisir b_n de façon à avoir

$$F^n(b_n) \to \Psi_\alpha(0) = 1$$

c'est-à-dire qu'on doit avoir

$$b_n \to x_0 .$$

Et, en vertu de (50) et du Lemme 2, nous pouvons faire ce choix en posant

(51) $$b_n = x_0 .$$

En vertu du Lemme 4 et de (51) la relation (47) est équivalente à la relation suivante: pour $n \to \infty$, on a

(52) $$n(1 - F(a_n x + x_0)) \to (-x)^\alpha.$$

Nous tirons tout d'abord de cette relation l'égalité $\lim_{n\to\infty} a_n = 0$. En effet, on a $a_n x + x_0 < x_0$ pour $x < 0$ et pour que le premier membre de (7) tende vers une limite finie il est nécessaire que l'égalité $\lim_{n\to\infty} (a_n x + x_0) = x_0$ soit vérifiée pour tout $x < 0$.

Supposons maintenant que $y \to -0$. Pour tout $y < 0$ suffisamment petit il est possible de trouver un n suffisamment grand pour avoir, soit

$$-a_n \leqq y \leqq -a_{n+1} ,$$

si $a_{n+1} \leqq a_n$, soit

$$-a_{n+1} \leqq y \leqq -a_n ,$$

si $a_n \leqq a_{n+1}$. Dans le premier cas nous voyons que

$$1 - F(-a_{n+1} + x_0) \leqq 1 - F(y + x_0) \leqq 1 - F(-a_n + x_0)$$

et que, pour tout $k > 0$,

$$1 - F(-a_{n+1}k + x_0) \leqq 1 - F(ky - x_0) \leqq 1 - F(-a_n k + x_0)$$

d'où il résulte que

$$\frac{1 - F(-a_{n+1} k + x_0)}{1 - F(-a_n + x_0)} \leqq \frac{1 - F(ky + x_0)}{1 - F(y + x_0)} \leqq \frac{1 - F(-a_n k + x_0)}{1 - F(-a_{n+1} + x_0)} .$$

Dans le second cas nous obtenons d'une façon analogue l'inégalité

$$\frac{1 - F(-a_n k + x_0)}{1 - F(-a_{n+1} + x_0)} \leqq \frac{1 - F(ky + x_0)}{1 - F(y + x_0)} \leqq \frac{1 - F(-a_{n+1} k + x_0)}{1 - F(-a_n + x_0)} .$$

Or, puisqu'en vertu de (52) dans les deux cas envisagés les membres extrêmes des inégalités obtenues tendent, pour $n \to \infty$, vers k^α, nous arrivons à la deuxième condition du théorème.

6. Le domaine d'attraction de la loi $\Lambda(x)$

Nous avons vu dans les paragraphes précédents que les lois $\Phi_\alpha(x)$ n'attirent que des fonctions de distribution pour lesquelles on a $F(x) < 1$ pour tous les x, et que les lois $\Psi_\alpha(x)$ n'attirent que des fonctions pour lesquelles on a $F(x_0) = 1$ pour une valeur finie de x_0 et $F(x_0 - \epsilon) < 1$ pour tout $\epsilon > 0$. Il est facile de voir que la loi $\Lambda(x)$ attire des fonctions des deux espèces envisagées. Donnons en des exemples.

Exemple 1. Soit

$$F(x) = \begin{cases} 0 & \text{pour} \quad x < 0 \\ 1 - e^{-x^\alpha} & \text{pour} \quad x > 0, \end{cases}$$

où $\alpha > 0$ est constant.

On trouve sans difficulté que, pour $n \to \infty$, on a

$$F^n(a_n x + b_n) \to e^{-e^{-x}},$$

où

$$a_n = \frac{1}{\alpha} (\log n)^{-\frac{\alpha-1}{\alpha}}, \qquad b_n = (\log n)^{1/\alpha}.$$

Exemple 2. Soit

$$F(x) = \begin{cases} 0 & \text{pour} \quad x \leqq 0 \\ 1 - e^{-\frac{x}{1-x}} & \text{pour} \quad 0 < x \leqq 1 \\ 1 & \text{pour} \quad x > 1. \end{cases}$$

On a aussi, pour $n \to \infty$,

$$F^n(a_n x + b_n) \to e^{-e^{-x}},$$

si l'on convient de poser

$$a_n = \log^{-2} n, \qquad b_n = \frac{\log n}{1 + \log n}.$$

Lemme 5. *Si, pour certains a_n et b_n et pour toutes les valeurs de x, on a*

(53) $$\lim_{n\to\infty} n(1 - F(a_n x + b_n)) = e^{-x},$$

alors

(54) $$\frac{a_n}{b_n} \to 0$$

pour $n \to \infty$.

DÉMONSTRATION: En effet, de (53) nous concluons que, pour toute valeur de x et pour $n \to \infty$

$$a_n x + b_n \to \infty \tag{55}$$

si $F(x) < 1$ pour toutes les valeurs de x; et que

$$a_n x + b_n \to x_0 , \tag{56}$$

si $F(x_0) = 1$ tandis que $F(x_0 - \epsilon) < 1$ pour tout $\epsilon > 0$. Soit $x = -A(A > 0)$; il est évident alors, en vertu de (55) et (56) que, dans le cas $x_0 > 0$, on a pour tout A et pour n suffisamment grand

$$a_n A < b_n ,$$

c'est-à-dire

$$\frac{a_n}{b_n} < \frac{1}{A} .$$

A étant arbitraire, la relation (54) en résulte.

Si $x_0 < 0$, nous tirons de (56), en y posant $x = 0$, pour $n \to \infty$ la relation

$$b_n \to x_0.$$

Or, de cette même relation (56), en y posant $x = 1$, nous tirons pour $n \to \infty$

$$a_n + b_n \to x_0 ,$$

d'où $a_n \to 0$ pour $n \to \infty$ et, par conséquent, $a_n/b_n \to 0$ pour $n \to \infty$. Si c'est lę cas $x_0 = 0$ qui se présente, alors, bien que $b_n \to x_0 = 0$ pour $n \to \infty$, nous avons $b_n < 0$ pour tout n suffisamment grand, puisque s'il n'en étsit pas ainsi nous aurions dû avoir, pour $x \geq 0$, $a_n x + b_n$ et cela nous conduirait à l'égalité

$$n(1 - F(a_n x + b_n)) = 0$$

ce qui est en contradiction avec l'hypothèse du lemme. En vertu de (53), pour n suffisamment grand et pour tout x nous avons

$$a_n x + b_n < 0;$$

donc, pour $x > 0$

$$\frac{a_n}{-b_n} < \frac{1}{x} .$$

Or, cette égalité ayant lieu pour toute valeur de x, la relation (54) en résulte.

THÉORÉME 6. *Pour qu'une fonction de distribution $F(x)$ appartienne au domaine d'attraction de la loi $\Lambda(x)$ il faut et il suffit que la relation*

$$\lim_{n\to\infty} n(1 + F(a_n x + b_n)) = e^{-x} \tag{57}$$

ait lieu pour toutes les valeurs x, où les constantes b_n sont definies comme les plus petites valeurs de x donnant lieu aux inégalités

$$F(x - 0) \leqq 1 - \frac{1}{n} \leqq F(x + 0) \tag{58}$$

et les constantes a_n sont les plus petites valeurs de x satisfaisant aux inégalités

$$F(x(1 - 0) + b_n) \leqq 1 - \frac{1}{ne} \leqq F(x(1 + 0) + b_n). \tag{59}$$

Démonstration: En vertu du Lemme 4, les conditions énoncées sont suffisantes pour qu'une fonction de distribution $F(x)$ appartienne au domaine d'attraction de la loi $\Lambda(x)$.

Inversement, si $F(x)$ appartient au domaine d'attraction de la loi $\Lambda(x)$, alors, en vertu du Lemme 4, on doit avoir pour un certain choix des α_n et β_n

$$n(1 - F(\alpha_n x + \beta_n)) \to e^{-x} \tag{60}$$

pour toutes les valeurs de x et pour $n \to \infty$.

Quel que soit $\epsilon > 0$, en vertu de (60), pour n suffisamment grand doivent avoir lieu les inégalités

$$n(1 - F(\alpha_n \epsilon + \beta_n)) + \eta < 1 < n(1 - F(-\alpha_n \epsilon + \beta_n)) - \eta$$

et

$$n(1 - F(\alpha_n(1 + \epsilon) + \beta_n)) + \eta < \frac{1}{\epsilon} < n(1 - F(\alpha_n(1 - \epsilon) + \beta_n)) - \eta$$

où $\eta = \frac{1}{2e}(1 - e^{-\epsilon})$.

De ces inégalités et de (58) et (59) nous tirons, pour tout ϵ, les inégalités

$$-\alpha_n \epsilon + \beta_n \leqq b_n \leqq \alpha_n \epsilon + \beta_n ,$$

$$\alpha_n(1 - \epsilon) + \beta_n \leqq a_n + b_n \leqq \alpha_n(1 + \epsilon) + \beta_n .$$

Puisque ces inégalités ont lieu pour tout $\epsilon > 0$, n étant suffisamment grand, nous pouvons choisir une suite $\epsilon_n > 0$ ($\epsilon_n \to 0$ pour $n \to \infty$) de façon à avoir

$$-\alpha_n \epsilon_n + \beta_n \leqq b_n \leqq \alpha_n \epsilon_n + \beta_n ,$$

$$\alpha_n(1 - \epsilon_n) + \beta_n \leqq a_n + b_n \leqq \alpha_n(1 + \epsilon_n) + \beta_n .$$

La première de ces inégalités nous fournit l'inégalité

$$\left| \frac{b_n - \beta_n}{\alpha_n} \right| \leqq \epsilon_n$$

et la seconde, en réunion avec l'inégalité obtenue à l'instant, nous conduit à l'inégalité

$$\left| \frac{a_n}{\alpha_n} - 1 \right| \leqq 2\epsilon_n .$$

Par conséquent, en vertu du Lemme 2, nous pouvons affirmer que toutes les fois que la relation (60) est vérifiée pour un certain choix des α_n et β_n, il en est de mêmde (57), le choix des a_n et b_n étant effectué conformément à (58) et (59). Le théorème est donc démontré.

THÉORÉME 7. *Pour qu'une fonction de distribution $F(x)$ appartienne au domaine d'attraction de la loi $\Lambda(x)$ il faut et il suffit qu'il existe une fonction continue $A(z)$ telle que $A(z) \to 0$ pour $z \to z_0 - 0$ et que, pour toute les valeurs de x*

$$(61) \qquad \lim_{z \to x_0 - 0} \frac{1 - F(z(1 + A(z)x))}{1 - F(z)} = e^{-x},$$

le nombre $x_0 \leqq +\infty$ étant determine par le relations $F(x_0) = 1$, $F(x) < 1$ pour $x < x_0$.

DÉMONSTRATION: Supposons d'abord que $F(x)$ appartient au domaine d'attraction de la loi $\Lambda(x)$. Alors pour un certain choix des constantes $a_n > 0$ et b_n nous avons

$$(62) \qquad \lim_{n \to \infty} n(1 - F(a_n x + b_n)) = e^{-x}$$

pour toutes les valeurs de x. Il en résulte que

$$(63) \qquad \lim_{n \to \infty} n(1 - F(b_n)) = 1.$$

Il est évident que l'égalite $a_n = 0$ ne peut avoir lieu que pour un nombre fini de valeurs de n; nous pouvons donc toujours supposer que $a_n > 0$ pour toutes les valeurs de n. Il resulte du Théoreme 6 que nous pouvous considézer les b_n comme fonctions non décroissantes de n.

Posons $A(b_n) = a_n/b_n$ pour toutes les valeurs de n et, pour $b_{n-1} \leqq z \leqq b_n$, définissons $A(z)$ de facon qu'elle soit fonction continue et monotone de z. Il est évident que pour tout $z < x_0$ suffisamment grand il est toujours possible de détermier un entier n tel que $b_{n-1} \leqq z \leqq b_n$. Il en résulte que

$$(64) \qquad 1 - F(b_n) \leqq 1 - F(z) \leqq 1 - F(b_{n-1}).$$

En vertu de la définition de la fonction $A(z)$ on doit avoir soit

$$\frac{a_{n-1}}{b_{n-1}} \leqq A(z) \leqq \frac{a_n}{b_n},$$

soit

$$\frac{a_n}{b_n} \leqq A(z) \leqq \frac{a_{n-1}}{b_{n-1}}.$$

Dans le premier cas nous voyons que pour $x > 0$

$$a_{n-1}x + b_{n-1} \leqq z(1 + A(z)x) \leqq a_n x + b_n,$$

et pour $x < 0$

$$a_n x + b_{n-1} \leqq z(1 + A(z)x) \leqq a_{n-1}x + b_n.$$

Donc, dans le premier cas, nous avons pour $x > 0$

(65) $$1 - F(a_n x + b_n) \leqq 1 - F(z(1 + A(z)x)) \leqq 1 - F(a_{n-1}x + b_{n-1})$$

et pour $x < 0$

(66) $$1 - F(a_{n-1}x + b_n) \leqq 1 - F(z(1 + A(z)x)) \leqq 1 - F(a_n x + b_{n-1}).$$

De (63), (64) et (65) nous tirons pour $x > 0$, les inégalités

$$(n - 1)[1 - F(a_n x + b_n)] \leqq \frac{1 - F(z(1 + A(z)x))}{1 - F(z)} \leqq n[1 - F(a_{n-1}x + b_{n-1})].$$

Or, ces inégalités, en vertu des Lemmes 3, 2, et de la relation (62), impliquent, pour $x > 0$, l'égalité (61).

D'une façon analogue nous obtenons (61) a partir de (66). Des raisonnements analogues nous conduisent de nouveau à (61) dans le cas non considéré $((a_n/b_n) \leqq A(z) \leqq (a_{n-1}/b_{n-1}))$. Cela achève la démonstration de la nécéssité de la condition (61), Quant à la démonstration de la suffisance de la condition du théorème remarquons tout d'abord qu'il résulte de (61) l'égalité

(67) $$\lim_{z \to x_0 - 0} \frac{1 - F(z + 0)}{1 - F(z)} = 1.$$

En effet, puisque pour tout $x > 0$ nous avons,

$$F(z(1 + A(x)z)) \geqq F(z + 0),$$

nous pouvons écrire

$$1 \geqq \frac{1 - F(z + 0)}{1 - F(z)} \geqq \frac{1 - F(z(1 + A(z)x))}{1 - F(z)}.$$

D'où

$$1 \geqq \overline{\lim_{z \to x_0 - 0}} \frac{1 - F(z + 0)}{1 - F(z)} \geqq \underline{\lim_{z \to x_0 - 0}} \frac{1 - F(z + 0)}{1 - F(z)} \geqq e^{-x}.$$

Ces inégalités ayant lieu pour toutes les valeurs de x, elles subsistent à la limite pour $x \to 0$.

Supposons maintenant que les conditions du théorème ont lieu et montrons que $F(x)$ appartient alors au domaine d'attraction de la loi $\Lambda(x)$. A cette fin définissons b_n comme étant la plus petite valeur de x donnant lieu à l'inégalité

$$1 - F(x + 0) \leqq \frac{1}{n} \leqq 1 - F(x - 0) = 1 - F(x).$$

Nous en obtenons que

$$\frac{1 - F(b_n(1 + A(b_n)x))}{1 - F(b_n - 0)} \leqq n(1 - F(b_n(1 + A(b_n)x)))$$
$$\leqq \frac{1 - F(b_n(1 - A(b_n)x))}{1 - F(b_n + 0)}.$$

En vertu de (61) et (67) nous voyons que

$$\lim_{n\to\infty} n(1 - F(b_n(1 + A(b_n)x)) = e^{-x}.$$

En posant $a_n = b_n A(b_n)$ nous obtenons (62). Le théorème est démontré.
Du théorème démontré résultent les propositions suivantes

Corollaire 1. *Supposons que la fonction de distribution $F(x)$ soit telle que $F(x) < 1$ pour toute valeur de x. Alors pour que la fonction $F(x)$ appartienne au domaine d'attraction de la loi $\Lambda(x)$ il faut que*

$$\frac{1 - F(ky)}{1 - F(y)} \to 0$$

pour tout $k > 0$ constant et pour $y \to \infty$, c'est-à-dire que la suite des maxima soit relativement stable.

Démonstration: Posons

$$\Phi_z(x) = \frac{1 - F(z(1 + A(z)x))}{1 - F(z)}; \tag{68}$$

les fonctions $\Phi_z(x)$ sont non croissantes par rapport à x. Nous avons eu plusieurs fois l'occasion de faire usage de cette remarque que si une suite de fonctions monotones converge en tout point vers une fonction continue, la convergence est uniforme. En vertu de la convergence uniforme de $\Phi_z(x)$ vers e^{-x}, x_z étant une suite tendant vers l'infini pour $z \to \infty$, nous devons avoir

$$\lim_{z\to\infty} \Phi_z(x_z) = \lim_{z\to\infty} e^{-x_z} = 0. \tag{69}$$

Prenons un $\alpha > 0$ et posons

$$x_z = \frac{\alpha}{A(z)}.$$

Par définition de la fonction $A(z)$ nous avons pour $z \to \infty$

$$\lim_{z\to\infty} A(z) = \infty.$$

Il résulte de (68) et (69) que

$$\lim_{z\to\infty} \frac{1 - F((1 + \alpha)z)}{1 - F(z)} = 0.$$

On voit facilement que la condition nécessaire que nous venons de trouver n'est nullement suffisante. Pour le montrer considérons une fonction de distribution définie de la façon suivante

$$F(x) = \begin{cases} 0 & \text{pour} \quad x < 0 \\ 1 - e^{-[x]} & \text{pour} \quad x > 0, \end{cases}$$

où $[x]$ désigne l'entier de x et démontrons qu'il est impossible de choisir les a_n et b_n de facon à avoir

$$\lim_{n\to\infty} n(1 - F(a_n x + b_n)) = e^{-x}$$

pour toutes les valeurs de x.

En d'autres termes montrons qu'il ne peut pas exister de constantes a_n et b_n donnant lieu à l'égalité

$$\lim_{n\to\infty} ne^{-[a_n x + b_n]} = e^{-x}$$

pour tous les x ou, ce qui est équivalent, l'égalité

$$\lim_{n\to\infty} (\log n - [a_n x + b_n]) = -x$$

pour toutes les valeurs de x, bien entendu. Considérons la suite partielle $n_k = [e^{k+0,5}]$, ou k est un entier. Nous avons, pour $k \to \infty$, $\log n_k - k \to 0, 5$; par suite, il ne peut pas exister de b_{n_k} telles que la relation

$$\lim_{k\to\infty} (\log n_k - [a_{n_k} \cdot 0 + b_{n_k}]) = \lim_{k\to\infty} (\log n_k - [b_{n_k}]) = 0$$

ait lieu. Il est évident que ceci démontre la proposition en question.

COROLLAIRE 2. *Soit $F(x)$ une fonction de distribution. S'il existe une suite $x_1 < x_2 < \cdots < x_k < \cdots$, $\lim_{k\to\infty} x_k = x_0$[4] ($x_0 \leqq +\infty$) telle que*

$$\frac{1 - F(x_k - 0)}{1 - F(x_k + 0)} \geqq 1 + \beta, \tag{70}$$

où la constante β est positive, la fonction $F(x)$ ne peut pas appartenir au domaine d'attraction de $\Lambda(x)$.

DÉMONSTRATION: L'inégalité (70) est incompatible avec l'égalité (67), et celle-ci résulte de (61).

REMARQUE. Il résulte des Théorèmes 4 et 5 que si la condition (70) a lieu, $F(x)$ ne peut pas appartenir aux domaines d'attraction des lois $\Phi_\alpha(x)$ et $\Psi_\alpha(x)$.

EXEMPLE. La loi de Poisson

$$F(x) = \begin{cases} 0 & \text{pour} \quad x < 0 \\ \displaystyle\sum_{0 \leqq k < x} e^{-\lambda} \cdot \frac{\lambda^k}{k!} & \text{pour} \quad x > 0 \end{cases}$$

n'est attirée vers aucune des lois limites. En effet, en posant $x_k = k$, nous voyons que

$$\frac{1 - F(k - 0)}{1 - F(k + 0)} = \frac{\displaystyle\sum_{s \geqq k} \frac{\lambda^s}{s!}}{\displaystyle\sum_{s > k} \frac{\lambda^s}{s!}} \geqq \frac{k + 1}{\lambda};$$

donc la condition (70) a lieu pour $k + 1 > \lambda$.

[4] Le nombre x_0 ayant le même sens qu'au Théorème 7.

Le théorème suivant donne un critère nécessaire et suffisant simple de convergence vers la loi $\Lambda(x)$ pour un choix particulier des constantes a_n.

THÉORÈM 8. *Pour que, pour un certain choix de la constante positive a et des contantes reelles b_n, la fonction de distribution $F(x)$ satisfasse à la relation*

$$F^n(ax + b_n) \to \Lambda(x) \tag{71}$$

pour $n \to \infty$, il faut et il suffit que l'on ait

$$\frac{1 - F(\log x)}{1 - F(\log kx)} \to k^\alpha, \qquad \text{pour } x \to \infty \tag{72}$$

pour toute valeur constante $k > 0$, où $\alpha a = 1$.

DÉMONSTRATION: Si la condition (71) a lieu nous avons, en vertu du lemme 5, l'inégalité $F(x) < 1$ pour toutes les valeurs de x et nous voyons que $b_n \to \infty$ pour $n \to \infty$.

Ensuite, il est évident que la détermination des conditions sous lesquelles (71) a lieu équivaut à la détermination des conditions pour lesquelles on a

$$\lim_{n \to \infty} F^n(x + b_n) = e^{-e^{-\alpha x}}. \tag{73}$$

Posons

$$x = \log z, \qquad b_n = \log \beta_n, \qquad F_1(z) = F(\log z).$$

Il est évident que $F_1(z)$ est une fonction de distribution. Dans ces conditions la détermination des conditions sous lesquelles on a (73) revient à la même question pour

$$F_1^n(\beta_n z) = F^n(\log \beta_n z) \to e^{-z^{-\alpha}}. \tag{74}$$

Les conditions nécessaires et suffisantes pour que la relation (74) ait lieu ont été trouvées au §3; elles sont équivalentes à la condition (72).

Un critère commode dans les applications pour savoir si la loi $F(x)$, vérifiant la condition $F(x) < 1$ pour toute valeur de x, appartient au domaine d'attraction de la loi $\Lambda(x)$, a été énoncé par Misès dans son travail cité. La condition de Misès consiste en ceci:

Soit $F(x)$ une fonction, admettant, pour tous les x supérieurs à une certaine valeur x_0, des dérivées des deux premiers ordres. Posons

$$f(x) = \frac{F'(x)}{1 - F(x)};$$

alors, si

$$\lim_{x \to \infty} \frac{d}{dx}\left[\frac{1}{f(x)}\right] = 0,$$

on a

$$\lim_{n \to \infty} F^n\left(x_n + \frac{x}{nF'(x_n)}\right) = e^{-e^{-x}},$$

où x_n est la plus petite des racines de l'équation $1 - F(x) = 1/n$. Une proposition analogue peut être démontrée pour une fonction ne satistaisant pas à la condition $F(x) < 1$ pour tous les x. S'il existe un x_0 tel que, pour tout $\epsilon > 0$,

$$F(x_0 - \epsilon) < 1, \qquad F(x_0) = 1$$

et que, à partir d'un $x < x_0$, la fonction $F(x)$ admet les dérivées première et seconde et que

$$\lim_{x \to x_0 - 0} \frac{d}{dx}\left[\frac{1}{f(x)}\right] = 0,$$

alors $F(x)$ appartient au domaine d'attraction de la loi $\Lambda(x)$.

MATHEMATICAL INSTITUTE OF THE
ACADEMY OF SCIENCES OF THE U. S. S. R.

RÉFÉRENCES

1. M. FRÉCHET. *Sur la loi de probabilité de l'écart maximum.* Annales de la société polonaise de Mathématiques, t. VI, p. 93, Cracovie, 1927.
2. R. A. FISHER AND L. H. C. TIPPETT. *Limiting forms of the frequency distribution of the largest or smallest member of a sample.* Proceedings Cambridge Philos. Soc. V, XXIV, part II, p.p. 180–190, 1928.
3. R. DE MISÈS. *La distribution de la plus grande de n valeurs.* Revue Mathématique de l'union interbalkanique, t. I, f. 1, p. 141–160, 1939.
4. A. KHINTCHINE. *Théorèmes limites pour les sommes des variables aléatoires indépendantes* (En russe), Moscou 1938.
5. B. GNEDENKO. *To the theory of the domains of attraction of stable laws,* Uchenye Zapiski of the Moscow State-University, t. 30, p.p. 61–81 (1939).
6. B. DE FINETTI. *Sulla legge di probabilità degli estremi,* Metron v. IX, No. 3–9, p.p. 127–138 (1932).

ANNALS OF MATHEMATICS
Vol. 44, No. 3, July, 1943

TRANSFORMATION GROUPS OF SPHERES

BY DEANE MONTGOMERY[1] AND HANS SAMELSON

(Received December 7, 1942)

I

1. The compact Lie group G is said to be a transformation group of the space W or to act on the space W if the following conditions are satisfied:

a) to every element g of G there is associated a homeomorphism $g(x)$ [x in W] of W onto itself.

b) if g_1 and g_2 are elements of G then $g_1[g_2(x)] = (g_1g_2)(x)$.

c) the point $g(x)$ depends continuously on the pair (g, x).

Conditions a) and b) imply that to the identity element of G is associated the identity homeomorphism. The group G is said to act transitively if in addition to a), b), and c) the following fourth condition is satisfied:

d) for any two points x and y of W there is an element g in G such that $g(x) = y$.

When d) is satisfied we say that W is a homogeneous space under G.

In this paper we take for W the n-dimensional sphere S^n and study the question of what compact connected Lie groups can act transitively and effectively, (see 2 a) below), on S^n. In I we prove a theorem on the structure of such a group which shows us that our main concern in the study of this problem is with simple groups. In II we study the question for simple groups using the Killing-Cartan classification, and we find that in general only those simple groups can be transitive and effective on S^n which are well known to be so. In III we use our methods to draw some conclusions about the structure of certain subgroups of the rotation group of the n-dimensional sphere which we denote by R_n. Otherwise expressed R_n is the group of orthogonal transformations of determinant 1 on $n + 1$ real variables.

2. We begin by noting some definitions and facts which will be of use in the course of the paper. All groups considered are compact Lie groups and we make the usual convention that finite groups are special cases of these. Subgroups are always taken as closed. For theorems on topological groups and Lie groups see [13] and for a general discussion of transformation groups see [18].

a) Let G act on W as above. Let H be any subgroup and let x be any point of W. The set of points of the form $h(x)$, h in H, is called the orbit of x under H and is denoted by $H(x)$. We see that H acts transitively on $H(x)$. Similarly if M is any subset of W then $H(M)$ denotes the set of all points of the form $h(m)$, h in H and m in M. The elements g of G for which the transformation $g(x)$ is the identity transformation of W form a normal subgroup G_0. If G' is an arbitrary normal subgroup of G contained in G_0 then the factor group

[1] Guggenheim Fellow.

G/G' acts in a natural way on W and G and G/G' generate the same orbits. The group G is said to act effectively if G_0 contains only the element e. We notice that G/G_0 always acts effectively.

b) Let H be any subgroup of G. The set of left cosets of H is made into a space (decomposition space) by means of the natural topology as follows: a set of cosets is an open set, if the points of G belonging to these cosets form an open set in G. This space is called the coset space and is denoted by G/H (in analogy with factor group); it is a manifold the dimension of which is the difference of the dimensions of G and H. The group G acts and acts transitively on G/H if to the element g we let correspond the transformation of G/H which sends the arbitrary coset bH into the coset gbH. The mapping which sends the element g of G into the element gH of G/H is called the projection of G onto G/H.

It is well known that homogeneous space and coset space are equivalent concepts, as we now indicate. Let W be a homogeneous space under G. Choosing arbitrarily a point x of W let H be the subgroup of those elements h in G which have x as fixed point, $h(x) = x$. For each y in W consider the set of those g's for which $g(x) = y$; this set is a left coset of H, and this correspondence between the points of W and G/H is a homeomorphism. The transformation $g(x)$ may therefore be considered as a transformation of G/H; it coincides there with the above introduced mapping $bH \rightarrow gbH$. We call H the associated subgroup. We shall frequently denote the group H here defined by the symbol G_x.

If we choose a different point x' to start with we simply have to perform an inner automorphism of G in order to find the associated subgroup. This is because the subgroup leaving x' fixed is $H' = g'Hg'^{-1}$, where g' is any element with the property $g'(x) = x'$.

The group G_0 (cf. a)) of elements which induce the identity transformation of W includes all normal subgroups of G contained in H; it is in fact the intersection of all groups conjugate to H. Therefore G is effective if and only if H contains no normal subgroup of G different from e.

c) The coset decomposition of G with respect to H is a special case of a fiber decomposition [4, 16]. We call a manifold M fibered with the fiber (-manifold) F, if the following holds: M is decomposed into sets homeomorphic with F, the "fibers"; every point of M is contained in one and only one fiber; and every fiber has a neighborhood in M which is homeomorphic to the topological product of F and a cell C, in such a way that a fiber is carried to a set which is the product of F and a point. The dimension of C is the difference of the dimensions of M and F. The decomposition space, which we get by considering the fibers as points, is a manifold (of the dimension of C).

d) The rank $r(G)$ of a compact connected Lie group (in this connection see [8]) is the dimension of a maximal Abelian subgroup of G, that is of an Abelian subgroup not contained in a larger Abelian subgroup (if G is not connected, we simply consider the component of the identity of G). Such a group is always a toral group, that is the direct product of a certain number of copies of the rotation group R_1 of the 1-sphere S^1. All maximal toral subgroups are conju-

gate, according to a fundamental theorem of Cartan. Each element of G lies on at least one maximal toral subgroup. The rank of a Lie group of positive dimension is always positive; we use in this paper "group of rank 0" as equivalent with "finite group." We note that if H is a normal subgroup of G and G/H the corresponding factor group, then:

$$r(G) = r(H) + r(G/H).$$

In particular, if G is the direct product of G_1 and G_2 (we denote this by $G = G_1 \times G_2$), then every maximal toral subgroup of G is of the form $T_1 \times T_2$, where T_i is a maximal toral subgroup of G_i $(i = 1, 2)$.

e) Let $G_1, \cdots, G_r$ be (compact connected) Lie groups and let N be a finite normal subgroup of the direct product $G = G_1 \times G_2 \times \cdots \times G_r$. We say that the factor group $G^* = G/N$ is essentially the product of $G_1, \cdots, G_r$. It is known that every compact connected Lie group is essentially the product of some simply connected simple groups and a toral group (see [13]).

3. We prove now a theorem on the structure of a group acting transitively on a sphere S^n. Let R_1 be the rotation group of the 1-sphere, and $\tilde{R}_2$ the simply connected covering group of the rotation group R_2 of the 2-sphere. The group $\tilde{R}_2$ may also, of course, be characterized as the group of quaternions of absolute value 1.

THEOREM I. *Let the compact connected Lie group G act transitively and effectively on S^n.*

a) *if n is even, then G is simple*

b) *if n is odd, then G is either simple or essentially the product of two simple groups G_1, G_2, where G_2 is either R_1 or $\tilde{R}_2$; and the subgroup of G corresponding to G_1 is transitive on S^n.*

In the course of the argument we shall also prove the following theorem:

THEOREM I′. *Let G_1 and G_2 be two compact connected Lie groups and let $G = (G_1 \times G_2)/N$ where N is a finite normal subgroup of $G_1 \times G_2$. If G is transitive on S^n then one of the two subgroups of G corresponding to G_1 and G_2 is transitive on S^n.*

In proving these two theorems we consider $G = G_1 \times G_2/N$ as given in the hypothesis of Theorem I′. We let $\bar{G} = G_1 \times G_2$ and we let $\bar{G}$ act in the natural way on S^n. We note that if G is effective then $\bar{G}$ is almost effective in the sense that only a finite number of its elements (in fact the elements of N) are the identity transformation of S^n. Let H be the associated subgroup of $\bar{G}$ which leaves fixed an arbitrary but definitely chosen point x of S^n. We shall find it convenient to identify the coset space $\bar{G}/H$ and S^n (see 2b)).

Theorem I is trivial for the case $n = 1$ as it is known that the only compact connected Lie group which can be effective on S^1 is R_1. Theorem I′ follows easily too because of the fact that for any x in S^1 $G_1(x)$ and $G_2(x)$ are sets which are either manifolds or contain only a single point. If both these sets contained only a single point, namely x, then $G(x)$ would also contain only the point x.

Hence either $G_1(x)$ or $G_2(x)$ is a manifold of positive dimension and must therefore coincide with S'.

In view of the above remarks we assume from now on that $n > 1$. The associated subgroup H is then connected, because of the fact that in this case S^n is simply connected [1]. We remark that we think of G_1 and G_2 as contained in $\bar{G}$; we notice that $g_1g_2 = g_2g_1$ (g_i in G_i) and $G_1 \cap G_2 = e$.

We shall also have need of the following statements about the ranks of $\bar{G}$ and H:

A) If n is even, then $r(\bar{G}) = r(H)$.

B) If n is odd, then $r(\bar{G}) = r(H) + 1$.

These statements follow from the fact that the rank $r(G)$ of an arbitrary compact connected Lie group G is equal to a certain homology invariant $1(G)$ which we do not define here [6] and to the fact that the statements A) and B) with $r(G)$ replaced by $1(G)$ are known to be true [14].

Proceeding now with the proof of Theorem I we first consider the case where n is even and to begin with we do not assume G to be effective. Let $h = h_1h_2$ (h_i in G_i) be any element of H. It is contained in a maximal toral subgroup T of H (cf. 2 d)). Because of 3 A) T is also a maximal toral subgroup of $\bar{G}$ and so is of the form $T_1 \times T_2$ where T_i is a maximal toral subgroup of G_i (cf. 2d)). Therefore the factors h_1 and h_2 of h are in T, and so in H. This means that H splits into a product $H_1 \times H_2$, where H_i is the intersection $H \cap G_i$ of H and G_i ($i = 1, 2$). But then the coset space $\bar{G}/H$ clearly decomposes into the topological product of the coset spaces G_1/H_1 and G_2/H_2:

$$\bar{G}/H = G_1/H_1 \times G_2/H_2$$

($=$ means "homeomorphic to"). On the other hand we have $\bar{G}/H = S^n$. Now if a sphere S^n is represented as the topological product of two manifolds, it follows from theorems on the homology of topological products that one of the two manifolds must be a point. We may suppose that G_2/H_2 is a point. This means that $G_2 = H_2$; a fortiori we have $G_2 \subset H$. Now G_2 being a normal subgroup of G we find according to 2 b) that the elements of G_2 induce the identity transformation of S^n. But then $G_1 = G/G_2$ must be transitive on S^n (cf. 2 a)); this proves Theorem I′ for even n.

Let us suppose now that G is effective. If G were not simple then $G = G_1 \times G_2/N$ as before. Then, as noted before, $\bar{G}$ is "almost" effective, and therefore H contains no infinite normal subgroup of $\bar{G}$. But G_2, being of positive dimension, is infinite. This contradiction shows that G must be simple, and theorem I a) is proved.

4. Now we let n be odd and for the present we do not require that G be effective. Again we have $\bar{G}/H = S^n$, and $\bar{G} = G_1 \times G_2$. We are unable to prove that H decomposes into the direct product of its intersections with G_1 and G_2. Therefore we consider the smallest subgroup Γ of $\bar{G}$, which contains H which decomposes into a direct product: $\Gamma = \Gamma_1 \times \Gamma_2$, where Γ_i is a

subgroup of G_i $(i = 1, 2)$. Obviously Γ_i is the image of H under the natural homomorphic mapping $g_1g_2 \to g_i$ of $\bar{G}$ onto G_i $(i = 1, 2;\ g_i$ in $G_i)$; it is clear from this that Γ_1 and Γ_2 are connected. Let H_i be the intersection $G_i \cap H$; we note that H_i is a normal subgroup of Γ_i $(i = 1, 2)$.

We show now that the coset spaces Γ/H, Γ_1/H_1, Γ_2/H_2 are homeomorphic. The easiest way to see this is by considering orbits. Let x be, as before, the point of S^n, which corresponds to the associated subgroup H. Clearly we have $\Gamma_i/H_i = \Gamma_i(x)$; for Γ_i acts transitively on $\Gamma_i(x)$, and the associated subgroup, if we choose the point x for its determination, is the intersection of Γ_i and H, and this is H_i. Similarly we have $\Gamma/H = \Gamma(x)$. Now Γ_1 and Γ_2 have the property that for each g_1 in Γ_1 there is a g_2 in Γ_2 such that $g_1(x) = g_2^{-1}(x)$, and the same with the indices interchanged; this follows immediately from the definition of Γ_i, because the condition $g_1(x) = g_2^{-1}(x)$ is equivalent with the condition g_1g_2 in H. This shows that $\Gamma_1(x) = \Gamma_2(x)$; and therefore, because $\Gamma(x) = \Gamma_1(\Gamma_2(x))$, both orbits are equal to $\Gamma(x)$. Moreover it is clear that this set is the intersection $G_1(x) \cap G_2(x)$; we denote it by F.

5. We determine now the structure of F. Consider the coset space Γ_1/H_1, homeomorphic with it. Since H_1 is a normal subgroup of Γ_1, it is a (compact connected Lie) group.

Suppose it is of rank 0 (cf. 2 d)). Being connected, it contains then only one element; this means $\Gamma_1 = H_1$, and therefore also $\Gamma_2 = H_2$. It follows that $H = H_1 \times H_2$. Hence the same argument as in 4 applies, and therefore theorem I′ follows in this case, that is under the assumption that the rank of Γ_1/H_1 is 0.

Suppose now that the rank of Γ_1/H_1 is positive. We show that in this case the rank is 1.

Let r, r_1, r_2 be the ranks of G, G_1, G_2; we have $r = r_1 + r_2$ (cf. 2 d)). The rank of H is $r - 1$ (cf. 3 B)). Let T be a maximal toral subgroup of H; its dimension is $r - 1$. It is contained in a maximal toral subgroup $\bar{T}$ of $\bar{G}$; and this $\bar{T}$ is of the form $T_1 \times T_2$, where T_i is a maximal toral subgroup of G_i and has dimension r_i (cf. nr. 2 d)). Now it is clear that the intersection $T \cap T_1$ is of dimension at least $r_1 - 1$. That means that the rank of $H_1 = H \cap G_1$ is at least $r_1 - 1$. On the other hand, Γ_1 being a subgroup of G_1, the rank of Γ_1 is at most r_1. Since H_1 is a normal subgroup of Γ_1, we have the equation $r(\Gamma_1) = r(\Gamma_1/H_1) + r(H_1)$. It follows that $r(\Gamma_1/H_1)$ is at most 1, and so, being positive, it is equal to 1. Because of the fact that Γ_1/H_1 is of rank 1, it is homeomorphic with one of the three following manifolds: the 1-sphere S^1, the 3-sphere S^3, the projective 3-space P^3 [15]; the same holds of course for F.

6. We have $\bar{G} \supset \Gamma = \Gamma_1 \times \Gamma_2 \supset H$, and the coset space Γ/H is homeomorphic with F (nr. 5). The decomposition of $\bar{G}$ into cosets of Γ induces, because of $\Gamma \supset H$, a decomposition of $\bar{G}/H$ into sets homeomorphic with Γ/H; one verifies easily because of the analytic nature of all imbeddings involved, that this is a

fiber decomposition of $\bar{G}/H$ with the fiber Γ/H (cf. nr. 2 c)). In symbols we have

$$\bar{G}/\Gamma = \bar{G}/H/\Gamma/H.$$

(This corresponds to a well known theorem in group theory—for the case of normal subgroups.) The right hand side of this formula means that $S^n(= \bar{G}/H)$ is fibered, and the fiber is $F(= \Gamma/H)$. In the left hand side we introduce the expressions $G_1 \times G_2$, $\Gamma_1 \times \Gamma_2$ for $\bar{G}$, Γ. As in nr. 4 we have obviously $\bar{G}/\Gamma = G_1/\Gamma_1 \times G_2/\Gamma_2$. Finally we get:

$$S^n/F = G_1/\Gamma_1 \times G_2/\Gamma_2 ;$$

the fiber space S^n/F is homeomorphic with the topological product of G_1/Γ_1 and G_2/Γ_2.

We know now that F is a homology-sphere (manifold with the Betti numbers of a sphere), of dimension d equal 1 or 3. Thus we are in a position to apply the theorems of Gysin [5], which give the result that the homology ring of S^n/F (with rational coefficients) has the following structure: it has a unit 1, corresponding to the fundamental cycle of the manifold; it has a certain element ξ of dimension $d + 1$ lower than that of 1; a certain power $(\xi)^m$ is 0-dimensional and different from 0; the elements $1, \xi, \xi^2 \cdots, (\xi)^m$ form a complete homology basis. It is now easy to see that this ring cannot occur as the homology ring of the topological product of two manifolds of positive dimensions. Therefore either G_1/Γ_1 or G_2/Γ_2 must reduce to a point; suppose this holds for G_2/Γ_2. Then we have $G_2 = \Gamma_2$.

7. From the fact just demonstrated it follows that G_1 acts transitively on S^n as we shall now show. The orbit $G_2(x)$ is equal to F, because $F = \Gamma_2(x)$ and $G_2 = \Gamma_2$. Consequently we have $G_2(x) \subset G_1(x)$, because $F = G_1(x) \cap G_2(x)$ (nr. 5). Therefore $G_1(G_2(x)) = G_1(x)$; but $G_1(G_2(x))$ is $\bar{G}(x)$, and $\bar{G}$ being transitive on S^n, this is all of S^n. Thus we have $G_1(x) = S^n$, and Theorem I′ is now proved completely.

Suppose now that G is effective. By definition H_2 is such that $H_2(x) = x$. If y is any other point of S^n there is an element g_1 in G_1 with the property that $y = g_1(x)$. Hence $g_1H_2g_1^{-1}(y) = y$. In view of the fact that elements of G_1 commute with those of G_2 we have $g_1H_2g_1^{-1} = H_2$. It has thus been shown that H_2 is a subgroup of $\bar{G}_0$ (nr. 2b)), that is H_2 leaves every element of S^n fixed. The group $\bar{G}$ is "almost" effective and consequently H_2 must be a finite group. But $\Gamma_2/H_2 = G_2/H_2$ is homeomorphic with Γ_1/H_1, and this means that G_2 itself is homeomorphic with a group of rank 1, that is with R_1 or R_2 or $\tilde{R}_2$. It is easy to see that it must then be isomorphic with one of these.

So far we have shown that, given any representation of G as essentially the product of two groups G_1 and G_2, then G_1 must be transitive over S^n and G_2 is isomorphic with R_1 or R_2 or $\tilde{R}_2$ (indices chosen properly). This clearly proves Theorem I b, if we show moreover that G_1 is simple.

Suppose it is not simple; let $G_1 = G' \times G''$ (possibly replacing G_1 by a finite covering group). We use now for G_1 the same argument we used for $\bar{G}$ and find that G' is transitive over S^n and that G'' is homeomorphic with R_1 or R_2 or $\tilde{R}_2$. We write now $\bar{G}$ as $G' \times (G'' \times G_2)$. The second factor $G'' \times G_2$ is not isomorphic with R_1 or R_2 or $\tilde{R}_2$; therefore, by what we have proved so far, it must be transitive on S^n. But this is impossible for $n > 3$, because, again by the results obtained so far, one of the factors G'', G_2 would have to act transitively, which is obviously impossible. The cases $n = 2, 3$ are treated easily for themselves. Theorem I is now proved completely.

It is worth pointing out that the possibilities mentioned in Theorem I may actually occur. There are groups which are essentially products of the type $G_1 \times R_1$ and $G_1 \times \tilde{R}_2$, where G_1 is simple, which act transitively and effectively on odd dimensional spheres. The group A_n (see for instance [14]) is defined as the group of all unitary matrices of determinant unity on $n + 1$ complex variables. The group, call it A'_n, of *all* unitary matrices on $n + 1$ complex variables is essentially the product of A_n and R_1. It is clear that A'_n is transitive on S^{2n+1} and it is also effective. Thus for every odd dimensional sphere S^{2n+1} of dimension greater than one there is a group of the type $G_1 \times R_1$ which is transitive on it.

A group of the type $G_1 \times \tilde{R}_2$ could not be effective on S^{4n+1} for $n = 1, 2, \cdots$. For if it were S^{4n+1} would be fibered by sets homeomorphic to S^3 or else by sets homeomorphic to P^3. These fibers are the orbits of the second factor. Either case is impossible as has been shown by Gysin [5].

However there is a group of the type of $G_1 \times \tilde{R}_2$ which is effective and transitive on S^{4n-1}, $n = 1, 2, \cdots$. Without giving details we shall merely mention that it may be obtained from C_n [see 14] in analogy with the way A'_n is obtained from A_n if we represent C_n by means of linear transformations on sets of n quaternions [19].

II

8. We now consider simple groups transitive on the sphere S^n. According to the Killing-Cartan classification every compact connected simple Lie group is locally isomorphic to one of the following simple groups: R_n, the rotation group of the n-sphere S^n (for $n \neq 1, 3$); A_n, the unitary unimodular group on $n + 1$ complex variables; C_n, the symplectic group on $2n$ complex variables; and five exceptional groups of dimensions 14, 52, 78, 133, 248 and ranks 2, 4, 6, 7, 8.

The dimension of R_n is $n(n + 1)/2$, that of A_n is $n(n + 2)$, and that of C_n is $n(2n + 1)$. The rank of R_n is $n/2$ for even n and $(n + 1)/2$ for odd n; the rank of A_n is n and the rank of C_n is n. The group R_n is transitive on S^n and A_n and C_n act in a natural way (as subgroups of the respective rotation groups) transitively on S^{2n+1} and S^{4n-1} respectively [14 p. 1126 ff.]. We shall speak of R_n, A_n, and C_n or any connected group locally isomorphic to them as the classical compact connected groups.

It will be necessary for us to use the homology properties of these groups and

we therefore list their homology rings. The symbol $R(M)$ denotes the homology ring with rational coefficients of the space M, the symbol $=$ denotes isomorphism, and x denotes as always the typological product. With these symbols, then, the following results are known (Brauer, Pontrjagin):

THEOREM A

a) $R(A_n) = R(S^3 \times S^5 \times \cdots \times S^{2n+1})$
b) $R(C_n) = R(S^3 \times S^7 \times \cdots \times S^{4n-1})$
c) $R(R_{2n}) = R(S^3 \times S^7 \times \cdots \times S^{4n-1})$
d) $R(R_{2n-1}) = R(S^3 \times S^7 \times \cdots \times S^{4n-5} \times S^{2n-1})$.

It is also known that if G_2 is the exceptional simple group of dimention 14 then

$$R(G_2) = R(S^3 \times S^{11}).$$

The rings of the four other exceptional simple groups are not known. If two compact connected Lie groups are locally isomorphic they have the same homology ring. This can be seen for example as follows. If N is a finite normal divisor of G, then the mapping taking G to G/N takes each homology group of G onto the corresponding group of G/N. Hence no Betti number can be increased by this mapping. On the other hand if r is the rank of G, the rank of G/N as also r, and since the sum of the Betti numbers of both groups is therefore 2^r [2] no Betti number can decrease. The Betti numbers of G and G/N are therefore the same, and it follows that the homology rings are isomorphic.

We shall use the following theorem on groups transitive on S^n which connects the homology properties of the group with those of the associated subgroup. This theorem has been proved by Samelson [14].

THEOREM B. *Let G be a compact connected Lie group which acts transitively on S^n and let H be the associated subgroup so that $G/H = S^n$.*

a) *if n is odd then $R(G) = R(H \times S^n)$ and H is not homologous to zero in G.*

b) *if n is even then $R(H) = R(\prod \times S^{n-1})$* where $\prod$ is a certain topological product of spheres of odd dimension and $R(G) = R(\prod \times S^{2n-1})$.

9. We have already seen that if n is even then any compact connected Lie group which acts effectively on S^n must be simple. We now examine in more detail the simple groups which can act in this way.

Before we begin it will be convenient to prove a lemma which will be of use in what follows. We recall [18, p. 202] that if G is effective on an n-dimensional orbit then the dimension of G is $\leqq n(n + 1)/2$.

LEMMA 1. *If a connected compact Lie group G of dimension $n(n + 1)/2$ is transitive and effective on a simply connected n-dimensional manifold M then M in an invariant metric is isometric to S^n and G is continuously isomorphic to R_n. Furthermore G_x is isomorphic to R_{n-1}.*

It is known [1] that we may introduce an invariant metric in M and [3] that in this metric M is of constant curvature. Hence [7] in view of the fact that M is compact and simply connected it must be isometric to the sphere S^n. The isometry T taking M to S^n carries G to a compact connected group TGT^{-1} of

rotations of S^n of dimension $n(n + 1)/2$. Therefore TGT^{-1} contains all rotations of S^n and we see that all the statements of the lemma are true.

THEOREM II. *If n is even, then except for a finite number of n's (the exceptional values of n being* $\leqq$ 114) *the only compact connected simple Lie group which can be transitive on* S^n *is locally isomorphic to* R_n.

The main step of the proof is contained in a theorem which we now state.

THEOREM II′. *If n is even the only compact classical group which can be transitive over* S^n *is locally isomorphic to* R_n.

According to theorem B b) the homology ring of G must be isomorphic to that of a space containing S^{2n-1} as a factor sphere. We know from Theorem A c) d) that if a group R_m has such a homology ring then m must be at least equal to n. On the other hand the dimension of G can not be greater than the dimension of R_n for if it were G could not be effective. Hence if one of the groups R_m is transitive on S^n, then m equals n.

If a group A_m has a homology ring of the kind we have observed G to have then A_m must have dimension greater than that of R_n and hence no A_m can be transitive on S^n. By similar considerations we observe that the only group C_m which could be transitive on S^n is $C_{n/2}$. This group can not be transitive on S^n as we see from Lemma 1. This concludes the proof of Theorem II′.

Theorem II in its general statement follows from Theorem II′. In order to obtain a limit on the dimensions of even dimensional spheres on which there are exceptions one must proceed to the direct consideration of the five exceptional simple groups. Here, although slightly sharper results can be derived, we are content with the one already stated. The dimension of the highest dimensional exceptional group G_{248} is 248. We know that if this is transitive on an even dimensional sphere S^n then

$$R(G_{248}) = R(S^{n_1} \times S^{n_2} \times \cdots \times S^{n_7} \times S^{2n-1})$$

where n_i ($i = 1, \cdots, 7$) is an odd integer. Because it is known [2] that the one dimensional Betti number of a simple group vanishes n_i is greater than or equal to 3. Hence

$$21 + 2^{n-1} \leqq 248$$

$$n \leqq 114.$$

10. We next consider the odd values of n and we consider separately the two possibilities $n \equiv 1$ (4), and $n \equiv 3$ (4).

THEOREM III. *Let* $n \equiv 1(4)$. *The only classical compact Lie* groups which act transitively on S^n are locally isomorphic to R_n and $A_{(n-1)/2}$.

It follows as before that with a possible finite number of exceptions these are the only simple Lie groups transitive on S^n with $n \equiv 1$ (4).

The proof of Theorem III is also based on the properties of homology rings.

According to Theorem B a), G must "contain" S^n as a factor sphere in its homology ring. Therefore it is clear at once that G cannot be C_m for any m or

R_m for any even m; and that the only possible R_m with m odd is R_n. Moreover it follows, that if G is a group A_m, then $m \geqq \frac{n-1}{2}$. Suppose $m > \frac{n-1}{2}$, then the ring of the associated subgroup H would be, according to Theorem B a), isomorphic with $R(S^3 \times S^5 \times \cdots \times S^{m-2} \times S^{m+2} \times \cdots \times S^{2m+1})$. This leaves for H only few possibilities: H could equal R_4 or C_2 or one of the exceptional groups $F_4E_6E_7E_8$, because no other group has a homology ring of this structure. If we take n sufficiently high, these possibilities are excluded and Theorem III is proved.

11. Next we wish to prove:

THEOREM IV. *Let $n \equiv 3(4)$; except for a finite number of n's the only simple groups transitive on S^n are R_n, $A_{(n-1)/2}$, $C_{(n+1)/4}$.* The proof of this is considerably more difficult than the proofs of the two preceding theorems.

Before beginning it is convenient to prove several lemmas which will be of use later. A stationary point of a group (or subset of a group) is a point (of the space on which the elements act) left fixed by every element of the group (or subset).

LEMMA 2. *Let M be a subset of the group of rotations R_n acting as it ordinarily does on S^n. Then $S(M)$, the set of stationary points of M, is a geometric sphere of dimension* $-1, 0, 1, \cdots$.

Let $\bar{S}(M)$ be the stationary points of M when we consider M as acting in E_{n+1}. Then $\bar{S}(M)$ is a linear substance of E_{n+1}. But

$$S(M) = \bar{S}(M) \cap S^n$$

and the conclusion follows.

LEMMA 3. *If H is a connected closed subgroup of R_n of dimension $n(n-1)/2$ then H is continuously isomorphic to R_{n-1} or $\tilde{R}_{n-1}$, the simply connected covering group of R_n.*

Let $R_n/H = M_n$ and let x be a point of M_n left fixed by H. The transformations of H act linearly on the space of tangent vectors to M_n at x. Hence H is a group of linear transformations of E_n and since H is compact we may assume that these linear transformations are orthogonal. We know that R_n has either no normal subgroups or at most a normal subgroup containing two elements. Therefore either R_n or R_n/Z when Z contains two elements is effective. If a group is effective, a subgroup is also. Hence H is either effective or H/Z is effective where Z contains two elements. Hence, in view of the dimension of H, either H or H/Z is R_{n-1} and hence H is either R_{n-1} or $\tilde{R}_{n-1}$.

LEMMA 4. *The group R_n contains no proper subgroup whose dimension is greater than dim R_{n-1}.*

Since R_n is simple, the group R_n has at most a finite set of elements leaving fixed all points of a coset space R_n/H. Hence dim $R_n/H \geqq n$ and consequently dim $H \leqq n(n+1)/2 - n = n(n-1)/2$.

The set of elements of R_n which leave fixed the unit point on the first k of the

$n + 1$ axes of E_{n+1} is isomorphic to R_{n-k}. This is one of many possible imbeddings of R_{n-k} in R_n. The particular subgroup picked out in this way we shall denote by Q_{n-k}. The subgroup which leaves invariant the first axis we denote by $\bar{Q}_{n-1}$. We see that $\bar{Q}^*_{n-1} = Q_{n-1}$ where $\bar{Q}^*_{n-1}$ is the identity component of $\bar{Q}_{n-1}$.

LEMMA 5. *If H is a proper closed subgroup of R_n which includes Q_{n-1}, then H is either Q_{n-1} or $\bar{Q}_{n-1}$.*

Since H can not have larger dimension than Q_{n-1}, H must consist of a finite number of cosets of Q_{n-1}. Consider the action of H on $S^n = R_n/Q_{n-1}$, and let p denote the unit point on the first axis so that Q_{n-1} leaves p fixed. The group H is in the normalizer of Q_{n-1}, and hence each coset $h\, Q_{n-1}$ takes x to a point left fixed by Q_{n-1}. Hence H can include at most two components and if there is a second one H must be $\bar{Q}_{n-1}$.

LEMMA 6. *Let R_m be imbedded in any way in R_n. Assume for every x of S^n, that R_{mx} (the subgroup of R_m leaving x fixed) is conjugate in R_m to one of the subgroups Q_{m-1} or $\bar{Q}_{m-1}$ of R_m. Then $m \leqq n/2$.*

Let F denote the set of x's such that R_{m_x} is conjugate to $\bar{Q}_{m-1}$ and let 0 denote the set of x's such that E_{m_x} is conjugate to Q_{m-1}. The set F is closed, 0 is open, both F and 0 are invariant under R_m, and

$$F + 0 = S^n.$$

In order to see that F is closed notice that if x is in F, then since R_{m_x} is conjugate to $\bar{Q}_{m-1}$, it must be true that $\bar{Q}_{m-1}$ has a stationary point in $R_m(x)$. That is, F is equal to the totality of orbits of R_m for points in S $(\bar{Q}_{m-1})$. We shall now show that F contains no points. Assuming that F does contain points we notice that $S(\bar{Q}_{m-1})$ must contain one and only one point on each orbit in F. That is $S(\bar{Q}_{m-1})$ is a cross section of F and F is the topological product of this cross section by an orbit of F. This is because we are dealing with a family of orbits. Each orbit of F is homeomorphic to m dimensional projective space P^m so we have shown that F is homeomorphic to the topological product of $S(\bar{Q}_{m-1})$ and P^m. In view of Lemma 2 it follows that $S(\bar{Q}_{m-1})$ is a sphere of some dimension.

We now define a homeomorphism T of period 2 taking S^n into itself whose set of fixed points is exactly F. We do this simply by leaving points of F fixed and by interchanging "diametral" points on orbits in 0, the concept of "diametral" points being defined on these orbits as a pair of points left fixed by precisely the same subgroup of R_m. We see that T is a homeomorphism as follows. Let x be any point of S^n. If y is near x then R_{m_y} is "near" R_{m_x}. This implies that the set of stationary points of R_{m_y} is near the set of stationary points of R_{m_x}. The existence of a T with these properties shows that F has the same homology groups mod 2 as a sphere of some dimension [17]. This is inconsistent with the above representation of F as a topological product of two manifolds unless F is vacuous. Hence we have shown that F is vacuous.

But then S^m is fibered (because of the analyticity of all the imbeddings involved) by sets homeomorphic to S^m, and so $m \leqq n/2$ [5].

LEMMA 7. *Let H be a connected group of dimension $n(n-1)/2$ imbedded in R_n. Then if $n \neq 3$, $n \neq 7$ H is conjugate to Q_{n-1}.*

We know in any case by Lemma 3 that H is isomorphic to either R_{n-1} or $\tilde{R}_{n-1}$. This means that we know the homology ring of H completely. We shall consider the action of H on the coset space $R_n/R_{n-1} \doteq S^n$, and we shall first show that H can not be transitive on S^n.

We first consider the case where $n = 2m$. In this case

$$R(H) = R(S^3 \times S^7 \times \cdots \times S^{4m-5} \times S^{2m-1})$$

and from Theorem B we see that H can not be transitive on S^{2m}.

We consider next the case where $n = 2m - 1$. In this case

$$R(H) = R(S^3 \times S^7 \times \cdots \times S^{4m-5}).$$

clearly if $2m - 1$ is of the form 5, 9, 13, $\cdots$ H can not be transitive. We therefore examine only the case where $2m - 1$ is of the form, 3, 7, $\cdots$, that is $2m - 1 = 4k - 1$, so that

$$\mathrm{R}(H) = R(S^3 \times S^7 \times \cdots \times S^{4k-1} \times \cdots \times S^{8k-1}).$$

Hence if $H/U = S_{4k-1}$

$$R(U) = (S^3 \times \cdots \times S^{4k-5} \times S^{4k+3} \times \cdots \times S^{8k-1}).$$

Now U is simple because only one S^3 appears on the right and it clearly can not be locally isomorphic to a classical group. Furthermore it can be checked, merely by enumerating possibilities that U can not be locally isomorphic to the last four exceptional groups. This leaves only one possibility, namely that H is R_6,

$$R(H) = R(S^3 \times S^7 \times S^{11})$$

$$R(U) = R(S^3 \times S^{11})$$

and $H/U \doteq S^7$.

We have now established that H is not transitive on S^n. Because of dimensional considerations it must have an $n - 1$ dimensional orbit. This is because of the fact that if H did not have an orbit of dimensions at least $n - 1$ it could not be effective and at the same time have dimension $n(n-1)/2$. In this connection see [18]. Hence all orbits are $n - 1$ dimensional, except for two orbits of lower dimension (see [11, and 18] and also the following lemma). However the only orbit of R_{n-1} or $\tilde{R}_{n-1}$ of lower dimension than $n - 1$ is a point. Let this point be x. Then H is in R_{m_x}. But R_{n_x} is conjugate to Q_{n-1}. This means that H is conjugate to a subgroup of Q_{n-1} which can only happen if H is conjugate to Q_{n-1}.

In connection with the following lemma we refer to [11] where a discussion of an analogous question is carried out.

LEMMA 8. *Let G be a subgroup of R_n which in its action on $R_n/R_{n-1} = S^n$ has an $(n-1)$-dimensional orbit. Then all orbits of G are $(n-1)$ dimensional*

except for two orbits of lower dimension. If one of these is not a point it carries a cycle linked with some cycle carried by the other.

It is known that all orbits with two exceptions are $n - 1$ dimensional and that the decomposition space of the orbits of G is an arc ab where the end points a and b correspond to orbits of lower dimension. The identity components of the groups G_x and G_y are conjugate if x and y are in $n - 1$ dimensional orbits [10, 11] and from this it can be seen that G_x and G_y are conjugate under the same circumstances [11]. This fact gives us a way of deforming sets from one $n - 1$ dimensional orbit to another, and in general a way of deforming a figure as long as we stay in the open set of $n - 1$ dimensional orbits [11]. This deformation can in fact be carried up to either of the exceptional orbits. These considerations together with the Alexander duality theorem prove the lemma.

We are now ready to proceed with the proof of Theorem IV.

For convenience put $m = (n + 1)/4$, so that $n = 4m - 1$. Leaving aside the exceptional groups (which means that we choose n high enough) we ask which of the classical groups can be transitive over S^n. From arguments like those above, namely by consideration of the homology ring of the associated subgroup, it follows that for n large enough, from the class A only $A(n - 1)/2$, and from the class C only C_m can be transitive. Similarly among the groups R_k with odd k only R_n can be transitive. Finally the only R_k with k even, which is not excluded by this consideration of the homology properties, is R_{2m}. Consequently what we have to prove is that R_{2m} can not act transitively on S^{4m-1}. Suppose it does. The associated subgroup H, having $R(S^3 \times S^7 \times \cdots \times S^{4m-5})$ as its homology ring according to Theorem B a), must be locally isomorphic with either R_{2m-2} or C_{m-1}.

a) Suppose H is locally isomorphic to C_{m-1}. We consider R_{2m} acting in the natural way on S^{2m}; then H, as a subgroup of R_{2m}, acts on S^{2m} too. By Theorem II′ it can not be transitive. On the other hand some orbit must be at least $(2m - 2)$-dimensional. For otherwise, since H is simple, all others would have to be points, and R_{2m} would not be effective. But Lemma 1 (Nr. 2) shows that H, being locally isomorphic with C_{m-1}, can have no $(2m - 2)$-dimensional orbit. Therefore H must have at least one orbit of dimension $2m - 1$.

As we have already observed it then follows that all orbits with the exception of two are $(2m - 1)$ -dimensional, and that the two exceptional orbits have lower dimensions. The orbit space, i.e. the decomposition space of the decomposition of S^{2m} into the orbits under H, is an arc, the end points of which correspond to the exceptional orbits. Now these two orbits can not, as stated above, be of dimension $2m - 2$. Therefore they must be of dimension 0; because a group acting effectively on a space of dimension k is of dimension at most $(k(k + 1))/2$; and C_{m-1} being simple, is effective or at least "almost" effective on any orbit not a point, and of dimension $((2m - 2)(2m - 1))/2$. This means that H has a stationary point, and so is contained in the associated subgroup of R_{2m}, that is in R_{2m-1}. But this is impossible, as we shall now see by considering the coset space R_{2m-1}/H. By considering H as acting on the line ele-

ments in the point of the space left fixed by it, we see that H is a subgroup of R_{2m-2}; but having the same dimension as R_{2m-2} it must coincide with R_{2m-2}, which contradicts the fact that it is locally isomorphic to C_{m-1}.

b) Suppose H is locally isomorphic with R_{2m-2}. Again we consider R_{2m} acting in the natural way on S^{2m}; and H acting on S^{2m} as a subgroup.

We know that H is not transitive on S^{2m}. We shall now show that no orbit of H can be $(2m - 1)$ dimensional. If some orbit of H is $(2m - 1)$ dimensional, then except for two orbits $H(x)$ and $H(y)$ which are of lower dimension all orbits are $(2m - 1)$ dimensional. If either $H(x)$ or $H(y)$ is a point then we see that H is conjugate to a subgroup of Q_{2m-1}. Furthermore it would have to be canonically imbedded in Q_{2m-1} (lemma 7) and $R_{2m/H}$ would be homeomorphic to R_{2m}/Q_{2m-2} but this space is not homeomorphic to S^{4m-1}. It is in fact homeomorphic to the space of all line elements on S^{2m}.

Since neither $H(x)$ nor $H(y)$ is a point they must both be $2m - 2$ dimensional. Then H_x, or at any rate its identity component, is isomorphic to R_{2m-3} (Lemma 1). Let z be any point of S^{2m} not in $H(x)$ nor in $H(y)$. If z is sufficiently near to x, H_z is isomorphic to a subgroup of H_x [11]. But $\dim H_z = \dim R_{2m-3} - 1$, and R_{2m-3} can contain no subgroup of this dimension. We have now completely shown that H can have no $2m - 1$ dimensional orbit. Incidentally we have pointed out that H can have no zero dimensional orbit.

We may now say that every orbit of H on $R_{2m}/R_{2m-1} = S^{2m}$ is $2m - 2$ dimensional. Before proceeding we pause to examine more carefully the structure of H. The group R_{2m-2} has no finite normal divisors. On the other hand the fundamental group of R_{2m-2} is cyclic of order two. Thus there is only one group locally isomorphic to R_{2m-2} and this group is the simply connected covering group $\tilde{R}_{2m-1}$ of R_{2m-2}. We may now say that H is either R_{2m-2} or $\tilde{R}_{2m-2}$ and we next wish to eliminate the latter possibility. The only $2m - 2$ dimensional orbits which the group $\tilde{R}_{2m-2}$ can have (Lemma 7) are S^n and P^n (projective n-space). The normal divisor of $\tilde{R}_{2m-2}$ contains the identity e and one other element which we shall call a. Since $\tilde{R}_{2m-2}$ is connected we see that if $\tilde{R}_{2m-2}$ is transitive on S^n or P^n then a must have a fixed point. But then a leaves every point of S^n or P^n fixed. Hence if H were $\tilde{R}_{2m-2}$ the element a would leave every point of every orbit fixed and so would leave every point of S^{2m} fixed which is contrary to our hypothesis. It follows that H is the group R_{2m-2} itself.

Hence (Lemma 7) for every x, H_x is conjugate to Q_{2m-3} (meaning the canonically imbedded subgroup of R_{2m-2}) or to $\bar{Q}_{2m-3}$ depending on whether $H(x)$ is homeomorphic to S^{2m-2} or P^{2m-2} respectively.

But then by Lemma 6, $2m - 2$ must be less than or equal to $m/2$. With this contradiction the proof of Theorem IV is concluded.

III

12. Theorem V. *Except for a finite number of n's, R_n has no subgroup H such that* $\dim H = \dim R_{n-1} - k$, $1 \leqq k \leqq n - 3$.

Proof. If such an H exists consider the action of H on $R_n/R_{n-1} = S^n$.

Since R_n is effective in its action on S^n, it follows that H is also. We assume that H is connected, which merely means, that if it is not connected, we choose the component of the identity.

In view of the results we have established we see that, with a finite number of exceptional n's, H can not be transitive in its action on $R_n/R_{n-1} = S^n$. However, on the basis of dimensional considerations, H must have some $(n - 1)$-dimensional orbit. By Lemma 8 we see that all orbits are $(n - 1)$ -dimensional except two which we shall denote by $H(x)$ and $H(y)$.

The group H can not be effective on $H(x)$ and $H(y)$ because their dimensions are too low. Hence there must be an invariant subgroup H_1 of positive dimension of H such that every point of $H(x)$ is stationary under H_1, and a similar H_2 associated with $H(y)$. We assume that H_1 and H_2 are connected, which, again, is merely a way of saying that if they are not connected we choose their identity components.

Since H_1 is invariant we see that if z is a stationary point of H_1 then every point of $H(z)$ is also a stationary point of H_1. A similar remark applies to H_2. Thus if z is in $S^n - (H(x) + H(y))$ z can not be a stationary point of H_1, for if it were the stationary points of H_1 would separate S^n which would imply that every point of S^n is a stationary point of H_1 [12] and that H was not effective. For the same reason H_2 can have no stationary points in $S^n - (H(x) + H(y))$.

Therefore $S(H_1)$, the set of stationary points of H_1 is either $H(x)$ or $H(x) + H(y)$. In the latter case, the set $S(H_1)$ is not connected and from the fact (Lemma 2) that $S(H_1)$ is a sphere we see in this case that both $H(x)$ and $H(y)$ are points. This implies that for some g, $gHg^{-1} \subset R_{n-1}$ which is impossible.

Consequently $S(H_1) = H(x)$ and by similar arguments $S(H_2) = H(y)$. Assume that (where $0 < 1_i < n - 1$)

$$\dim H(x) = 1_1 \quad \text{and} \quad \dim H(y) = 1_2 .$$

Lemma 2 tells us that $H(x)$ and $H(y)$ are respectively 1_1-dimensional and 1_2-dimensional spheres.

At this point we shall need to recall (Lemma 8) that some cycle in $H(x)$ must link some cycle in $H(y)$.

From this it follows that the linked cycles are actually the basic cycles of the homology spheres and that

$$1_1 + 1_2 = n - 1.$$

We notice now that no element of H_1 can leave all of $H(y)$ fixed. The existence of such an element would enable us to find a subgroup of H_1 which has $H(y)$ as part of its stationary set. Such a group also has $H(x)$ as part of its stationary set. The stationary set, being a sphere including these two linked spheres, must coincide with S^n which is impossible. That is, H_1 is effective on $H(y)$ and similarly H_2 is effective on $H(x)$.

From the choice of H_1 we know that

$$\dim H/H_1 \leqq \frac{l_1(l_1 + 1)}{2}$$

$$\begin{aligned}\dim H_1 &\geqq \dim H - \frac{l_1(l_1 + 1)}{2} \\ &= \dim R_{n-1} - k - \frac{l_1(l_1 + 1)}{2} \\ &= \frac{(n - 1)(n) - 2k - l_1(l_1 + 1)}{2}.\end{aligned}$$

But

$$l_1 = n - l_2 - 1$$

and hence

$$\begin{aligned}\dim H_1 &\geqq \frac{(n - 1)(n) - (n - l_2 - 1)(n - l_2) - 2k}{2} \\ &= \frac{l_2(2n - l_2 - 1) - 2k}{2} \\ &= \frac{l_2(l_2 + 1) + l_2(2n - 2l_2 - 2) - 2k}{2}.\end{aligned}$$

At this point we remark that we assume $n \geqq 2$. We also assume as we may that $l_2 \leqq n/2$. By hypothesis, we know that $k \leqq n - 3$. Now

$$l_2(2n - 2l_2 - 2) - 2k \geqq l_2(2n - 2l_2 - 2) - 2n + 6.$$

This is clearly positive when $l_2 = 1$. Using the fact that $l_2 \leqq n/2$, we have

$$\begin{aligned}l_2(2n - 2l_2 - 2) - 2n + 6 &\geqq l_2(2n - n - 2) - 2n + 6 \\ &= (l_2 - 2)(n - 2) + 2\end{aligned}$$

which is clearly positive when $l_2 \geqq 2$. Hence we may always conclude that

$$\dim H_1 < \frac{l_1(l_2 + 1)}{2}.$$

But we proved above that H_1 is effective on $H(y)$ which is l_2 dimensional. Therefore

$$\dim H_1 \leqq \frac{l_2(l_2 + 1)}{2}.$$

These two contradictory facts show that no subgroup H of the kind described in the theorem exists.

From this theorem we see that except for a finite number of n's R_n can have no orbit $H(x)$ whose dimension satisfies the following inequality

$$n < \dim H(x) \leqq 2n - 3.$$

We also point out the following theorem which is a corollary of our results.

THEOREM VI. *Let H be a subgroup of R_n which is isomorphic to R_k, $k > n/2 + 2$. Then with the possible exception of a finite number of n's H is codjugate to Q_k.*

Since H is a subgroup of R_n we may consider the action of H on $R_n/R_{n-1} = S_n$. By the preceding theorem H can have no orbit $H(x)$ such that

$$k < \dim H(x) \leqq n.$$

Furthermore because of a previous lemma not all orbits can have dimension k. Hence H leaves some point stationary which implies that H is conjugate to a subgroup of R_{2n-1}. By a finite number of applications of our process we obtain the desired conclusion.

SMITH COLLEGE AND
UNIVERSITY OF WYOMING

BIBLIOGRAPHY

1. CARTAN, *La Theorie des Groupes Finis et Continus et l'Analysis Situs*, Memorial des Sciences Mathematiques, no. 42.
2. CARTAN, *La Topologie des Groupes de Lie*, Actualites Scient et Industr. 358; also in L'Enseignement Math. vol. 35 (1936), pp. 177–200.
3. EISENHART, *Continuous Groups of Transformations*, Princeton, 1933.
4. FELDBAU, *Sur la classification des espaces fibres*, C. R. Acad. Sci., Paris, vol. 208 (1939), pp. 1621–1623.
5. GYSIN, *Zur Homologietheorie der Abbildungen und Faserungen von Mannigfaltigkeiten*, Commentari Math. Helv. vol. 14 (1941–1942), pp. 61–122.
6. HOPF, *Über die Topologie der Gruppen-Mannigfaltigheiten und ihrer Verallgemeinerungen*, Annals of Math., vol. 42 (1941), pp. 22–52.
7. HOPF, *Zum Clifford-Kleinschen Raumproblem*, Math. Annalen, vol. 95 (1926), pp. 313–339.
8. HOPF, *Über den Rang geschlossener Liescher Gruppen*, Comment. Math. Helv. vol. 13 (1940), pp. 119–143.
9. MONTGOMERY AND ZIPPIN, *Topological Transformation Groups I*, Annals of Math., vol. 41 (1940), pp. 778–791.
10. MONTGOMERY AND ZIPPIN, *A Theorem on Lie Groups*, Bull. Am. Math. Soc., vol. 48 (1942), pp. 448–452.
11. MONTGOMERY AND ZIPPIN, *A Class of Transformation Groups in E_n*, American Jour. of Math. To appear shortly.
12. NEWMAN, *A Theorem on Periodic Transformations of Spaces*, Quarterly Jour. of Math., vol. 2 (1931), pp. 1–8.
13. PONTRJAGIN, *Topological Groups*, Princeton, 1939.
14. SAMELSON, *Beiträge zur Topologie der Gruppen-Mannigfaltigkeiten*, Annals of Math. vol. 42 (1941), pp. 1091–1137.
15. SAMELSON, *Über die Sphären, die als Gruppenräume auftreten*, Comment. Math. Helv. vol. 13 (1940), pp. 144–155.
16. SEIFERT, *Topologie Dreidimensionaler Gefaserten Raume*, Acta Mathematica, vol. 60 (1933), pp. 147–238.
17. SMITH, *Transformations of Finite Period*, Annals of Math., vol. 39 (1938), pp. 127–164.
18. ZIPPIN, *Transformation Groups*, Lectures in Topology, Michigan (1941), pp. 191–221.
19. JACOBSON, *Semi-Simple Lie Groups in the Large*, Annals of Math., vol. 40 (1939), pp. 755–763.

Annals of Mathematics
Vol. 44, No. 3, July, 1943

EXTENSIONS OF TOPOLOGICAL SPACES[1]

By S. Fomin

(Received February 8, 1943)

Introduction

An extension of a given topological space R is by definition any topological space containing R as an everywhere dense subset. The first essential results about extensions of topological spaces were obtained by A. Tychonoff [1], who proved that every completely regular space can be immersed into a bicompact one. An essential strengthening of this result was given by Čech [2]. He has shown the existence for every completely regular space of an universal bicompact extension βR, which is characterized by the following property: whatever be the bicompact extension bR of the space R, there exists a continuous mapping of βR on bR, leaving all points of R fixed. Another construction of the same extension was given by P. Alexandroff [3] in 1939. The method, which was applied in the paper of P. Alexandroff seems to us to be the most natural for the questions of the theory of extensions, and, slightly modified, it enables us to obtain from one and the same general considerations all facts, known in this domain (including the results of Stone on the H-closed extensions), as well as some new results, which, in our opinion, are of some interest. A knowledge of the paper of Čech is not presupposed in what follows.

In §§1–2 of the present paper we give a general construction of an arbitrary Hausdorff extension of a Hausdorff topological space, and introduce some notions, which are essential for the sequel.

In §3 we consider H-closed extensions of a Hausdorff space.

§4 contains a generalization of the notion of continuity which is applied in §5 to prove the existence, for every Hausdorff space, of an universal H-closed extension analogous to the Čech extension.

In §6 the theory of bicompact extensions is given.

1. A general method of construction of extensions of topological spaces

In the course of the whole paper we shall confine ourselves to the consideration of Hausdorff topological spaces and their Hausdorff extensions. In §§6 and 7 we shall assume even stronger separability axioms. In general, if the contrary is not explicitly stated, by "space" we shall always understand a Hausdorff topological space.

Let R be an arbitrary space, and $\mathfrak{G}$—one of its bases. Let $\mathfrak{S}$ be an aggregate of centered[2] systems $\mathfrak{s}$, composed of the elements of this basis; we suppose that $\mathfrak{S}$ satisfies the following conditions:

[1] An exposition of the main results of this paper has been given without proofs in my Note published under the same title in the Comptes Rendus de l'Acad. des Sciences de l'URSS 32 (1941), 114–116.

[2] We call the system $\mathfrak{s}$ of non-vacuous sets A centered, if the intersection of any two sets from $\mathfrak{s}$ belongs to $\mathfrak{s}$. In accordance with definition it is supposed that the considered basis $\mathfrak{G}$ possesses the property that the intersection of any two elements from $\mathfrak{G}$ also belongs to $\mathfrak{G}$.

1. For every point $r \in R$ there exists a centered system $\mathfrak{s}_r \in \mathfrak{S}$ namely, the system of all neighborhoods of the point r, belonging to the basis $\mathfrak{G}$.

2. If $\mathfrak{s}_1$ and $\mathfrak{s}_2$ are two different systems from $\mathfrak{S}$, then there exist open sets Γ_1 and Γ_2 from $\mathfrak{G}$ such that $\Gamma_1 \in \mathfrak{s}_1$, $\Gamma_2 \in \mathfrak{s}_2$ and $\Gamma_1 \cap \Gamma_2 = 0$.

Every set Γ, entering into the system $\mathfrak{s}$, will be called a coordinate of this system. Thus $\Gamma \in \mathfrak{s}$ means that "Γ is a coordinate of the system $\mathfrak{s}$."

We now introduce in the set $\mathfrak{S}$ a topology, by defining the system of neighborhoods of an arbitrary point $\mathfrak{s} \in \mathfrak{S}$ in the following way:[3] if $\Gamma \in \mathfrak{s}_0$: then the neighborhood $U_\Gamma(\mathfrak{s}_0)$ consists of all points $\mathfrak{s}$, such that $\Gamma \in \mathfrak{s}$.

It is easily seen that $\mathfrak{S}$ is actually a topological space. From the condition 2) follows the existence of non-intersecting neighborhoods of any two points of $\mathfrak{S}$. Thus $\mathfrak{S}$ is a Hausdorff space.

To every point $r \in R$ may be correlated a centered system which is the system of neighborhoods of this point. It is evident that this is a one-to-one correspondence. It is also continuous in both directions. In fact, let the point $r_0 \in R$ and G_0 be its arbitrary neighborhood, belonging to the basis $\mathfrak{G}$; then $r \in G_0$ if and only if $\mathfrak{s}_r \in U_{G_0}$.

The image of the space R is a set, everywhere dense in $\mathfrak{S}$; and it follows from condition 1) that every set U_Γ contains a certain point $\mathfrak{s}_r$. Thus $\mathfrak{S}$ contains an everywhere dense subset, homeomorphic to R, i.e. $\mathfrak{S}$ *is an extension of the space* R.

On the other hand, every extension X of the space R may be represented as a space of an aggregate of centered systems of open sets, satisfying the conditions 1–2. To this end it is sufficient to correlate to every point $x \in X$ the centered system $\{U(x) \cap R\}$, where $\{U(x)\}$ is the system of all neighborhoods of the point $x \in X$, and to introduce in this set of centered systems a topology in the way indicated above. That the conditions 1 and 2 are satisfied, is obvious.

Let us now formulate the obtained result:

THEOREM 1. *Every set of centered systems* $\mathfrak{S}$, *satisfying the conditions* 1 *and* 2, *is an extension of the original space* R. *Conversely, every extension of the space* R *can be represented as a space of a set of centered systems, satisfying the conditions* 1 *and* 2.

Choosing in an appropriate way the set $\mathfrak{S}$ of centered systems $\mathfrak{s}$, we may, using the general method exposed above, construct different extensions of the original space possessing some a priori given properties.

2. Some definitions. The spaces $\sigma_{\mathfrak{G}}R$, $\alpha_{\mathfrak{G}}R$ and $\alpha'_{\mathfrak{G}}R$

Consider some special types of centered systems, which shall play a fundamental rôle in the sequel. Let a certain basis $\mathfrak{G}$ be chosen in the space R. We introduce the following definitions.

A centered system $\{G\}$ of sets from $\mathfrak{G}$ is called a Hausdorff centered system if, for every point $x \in R$ which does not enter into a certain set G of the system $\{g\}$, there exists another set G', also belonging to the system $\{G\}$, whose closure does not contain the point x.

[3] This topology is introduced along the lines of P. Alexandroff [3].

A Hausdorff centered system is called a *Hausdorff end*, if it is *maximal*, i.e. if it is not a true subsystem of any Hausdorff centered system composed of elements of the basis $\mathfrak{g}$.

We shall require the following notions, introduced by P. Alexandroff [3].

The open set Γ' is said to be completely regularly enclosed in the open set Γ, if $\bar{\Gamma}' \subseteq \Gamma$ and if to every rational number t of the segment $[0, 1]$ may be correlated an open set $\Gamma_t \in \mathfrak{G}$ in such a way that from $t_1 < t_2$ follows $\bar{\Gamma}_{t_1} \subseteq \Gamma_{t_2}$ and $\Gamma' = \Gamma_0$, $\Gamma = \Gamma_1$.

Γ' is regularly enclosed in Γ, if $\bar{\Gamma}' \subseteq \Gamma$.

A centered system $\{G\}$ of open sets from $\mathfrak{G}$ is called regular (completely regular), if every set of the system $\{\mathfrak{G}\}$ regularly (completely regularly) contains a set of the same system enclosed in it.

A regular centered system is called a regular end, if it is not a subsystem of any regular centered system (composed of elements of the basis $\mathfrak{G}$).

A completely regular end is defined similarly.

We define now the topological spaces $\sigma_{\mathfrak{g}}R$, $\alpha_{\mathfrak{g}}R$ and $\alpha'_{\mathfrak{g}}R$, whose points are, respectively, Hausdorff, regular and completely regular ends $\xi = \{G\}$, composed of sets $G \in \mathfrak{g}$. The topology in these spaces is introduced in the way indicated in § 1.

The space $\sigma_{\mathfrak{g}}R$ shall be considered for an arbitrary Hausdorff space R, and $\alpha_{\mathfrak{g}}R$ and $\alpha'_{\mathfrak{g}}R$—only for a regular R and completely regular R, respectively. If $\mathfrak{G}$ is the system of *all* open sets of R, the spaces $\sigma_{\mathfrak{g}}R$, $\alpha_{\mathfrak{g}}R$, $\alpha'_{\mathfrak{g}}R$ are denoted by σR, αR, $\alpha' R$. The spaces αR and $\alpha' R$ were first constructed by Alexandroff [3].

LEMMA 1. *Denote by* $\{U(x)\}$ *the aggregate of all elements of the basis* $\mathfrak{G}$ *of the space* R, *containing a given point* x *of this space.* $\{U(x)\}$ *is a Hausdorff end,* $\{U(x)\}$ *is a regular (completely regular) end, if* R *is regular (completely regular).*

PROOF. We lead the proof for a Hausdorff space R. A centered system $\{U(x)\}$ is a Hausdorff system, since if $x' \bar{\in} U^{(1)}(x)$, then $x' \neq x$, and there exist non-intersecting neighborhoods $V(x')$ and $U^{(2)}(x) \in \{U(x)\}$. Then $x' \bar{\in} \overline{U^{(2)}(x)}$. It remains to prove that $\{U(x)\}$ is not contained in any larger Hausdorff centered system. Let us assume the contrary. Let $\{U\}$ be a Hausdorff centered system containing $\{U(x)\}$, and let its element be $U_0 \bar{\in} \{U(x)\}$. Then U_0 does not contain x, and consequently, there exists a $U_1 \in \{U\}$ such that $x \bar{\in} \bar{U}_1$. But then there exists a $U(x)$ not intersecting with U_1, which contradicts the fact that $\{U\}$ is centered.

The other assertions of the lemma are proved similarly.

We have thus proved that the spaces $\sigma_{\mathfrak{G}}R$, $\alpha_{\mathfrak{G}}R$ and $\alpha'_{\mathfrak{G}}R$ satisfy (in the case, when for R the corresponding separability axioms hold) condition 1), §1. It is easily seen that they also satisfy condition 2), since two Hausdorff (regular, completely regular) ends coincide, if each coordinate of the one end has a non-vacuous intersection with each coordinate of the other end.

Thus from the results of §1 follows the

THEOREM 2. *The spaces* $\sigma_{\mathfrak{G}}R$, $\alpha_{\mathfrak{G}}R$ *and* $\alpha'_{\mathfrak{G}}R$ *are Hausdorff extensions of the space* R.

3. *H*-closed extensions

DEFINITION. *A basis $\mathfrak{G}$ of the space R is called algebraically closed, if from $G \in \mathfrak{G}$ follows that also $R - \bar{G} \in \mathfrak{G}$.*

The following theorem will be essential in the sequel.

THEOREM 3. *If the basis $\mathfrak{G}$ is algebraically closed, then the space $\sigma_{\mathfrak{G}} R$ is H-closed*[4].

The proof of this theorem is based on the following criterion of H-closedness of a topological space.

THEOREM 4. *Let $\mathfrak{G}$ be an arbitrary algebraically closed basis of the space R. This space is H-closed if and only if every Hausdorff centered system, composed of elements of the basis $\mathfrak{G}$, has a non-void intersection.*

THE PROOF OF THEOREM 4

The condition is necessary. If there exists a Hausdorff centered system ξ, with vacuous intersection, then ξ may be adjoined to our space R as a new (non-isolated) point, taking for the neighborhoods of ξ the sets $\xi \cup G$, where G is an arbitrary coordinate of the system ξ. The space $R \cup \xi$ is a Hausdorff space, since for an arbitrary $x \in R$ there exists a coordinate G of ξ not containing x, and so, in virtue of the fact that the system ξ is a Hausdorff system, there exists another coordinate G' of ξ, whose closure does not contain x. Consequently x and ξ possess non-intersecting neighborhoods. The space R is not H-closed.

The condition is sufficient. Let the space R be not H-closed. Let us adjoin to R a new (non-isolated) point ξ, and consider the aggregate $\{U(\xi)\}$ of all neighborhoods of the point ξ, such that $R \cap U(\xi) \in \mathfrak{G}$.

It is easily seen that the sets $R \cap U(\xi)$ form a Hausdorff centered system.

Let x be an arbitrary point of the space R. Since the space $R \cup \xi$ is a Hausdorff space, there exists a neighborhood $V(x)$ (belonging to the basis $\mathfrak{G}$), whose closure (in $R \cup \xi$) does not contain the point ξ and therefore coincides with the closure in R. Then, in virtue of the algebraical closedness of the basis $\mathfrak{G}$, the set $R - \overline{V(x)}$ belongs to $\mathfrak{G}$. Now the set $(R \cup \xi) - \overline{V(x)}$ contains ξ, and consequently, the set

$$R - \overline{V(x)} = R \cap ((R \cup \xi) - \overline{V(x)})$$

is one of the coordinates of the considered centered system. The point x is not contained in this coordinate, and since x was chosen arbitrarily, the intersection of all coordinates of the considered system is void. Thus the sufficiency of the condition is also proved.

LEMMA 2. *If $\mathfrak{G}$ is an algebraically closed basis, and $\sigma_{\mathfrak{G}} R$ is the space corresponding to this basis, then*

$$\sigma_{\mathfrak{G}} R - \bar{U}_G = U_{R-\bar{G}}$$

PROOF. Let the Hausdorff end $\{\Gamma_\alpha\} \in \sigma_{\mathfrak{G}} R - \bar{U}_G$; then there exists a neighborhood U_Γ such that $U_\Gamma \cap U_G = 0$. But this means that $\Gamma \cap G = 0$ as well,

[4] The space R is called H-closed, if it is closed in every Hausdorff space, containing it.

and then we have also $\Gamma \cap \bar{G} = 0$, i.e. $\Gamma \subseteq R - \bar{G}$ and, consequently, $\{\Gamma_\alpha\} \in U_{R-\bar{G}}$. Conversely, let $\{\Gamma_\alpha\} \in \bar{U}_G$. Then $\Gamma_\alpha \cap G \neq 0$ for every Γ_α. If now $\{\Gamma_\alpha\} \in U_{R-\bar{G}}$, then $\{\Gamma_\alpha\} \ni (R - \bar{G})$, but this is impossible, since $(R - \bar{G}) \cap G = 0$.

From this lemma immediately follows:

COROLLARY. *If $\mathfrak{G}$ is an algebraically closed basis of the space R, then $\{U_G, G \in \mathfrak{G}\}$ is an algebraically closed basis of $\sigma_{\mathfrak{G}}R$.*

PROOF OF THEOREM 3. Let $\{U_{G_\alpha}\}$ be an arbitrary Hausdorff centered system composed of elements of the basis $\{U_G, G \in \mathfrak{G}\}$. Let us show that $\prod_\alpha U_{G_\alpha} = 0$. It is easily seen that the sets G_α form themselves a Hausdorff centered system $\{G_\alpha\}$. Complete the system up to a maximal. Then we obtain a certain Hausdorff end, i.e. a certain point ξ of the space $\sigma_{\mathfrak{G}}R$. It is obvious that every open set U_{G_α} contains the point ξ, i.e. $\prod_\alpha U_{G_\alpha} \neq 0$. By Theorem 4, the space $\sigma_{\mathfrak{G}}R$ is H-closed.

4. θ-continuous mappings

In studying non regular Hausdorff spaces the following definition is very useful.

DEFINITION. *Let X and Y be two topological spaces. A one-valued mapping f of X into Y is called θ-continuous at the point $x_0 \in X$ if for every neighborhood $U(y_0)$ of the point $y_0 = f(x_0)$ there exists a neighborhood $U(x_0)$ such that $f(\overline{U(x_0)}) \subseteq \overline{U(y_0)}$.*

A (1-1)-mapping of X onto Y which is θ-continuous in both directions is called a *θ-homeomorphism* between X and Y.

A θ-continuous mapping of a space X into a *regular* space Y is continuous in the ordinary sense. In all cases a continuous mapping is θ-continuous, the converse being for irregular Y in general not true.

Example. Let R_1 be the segment [0, 1] with its ordinary topology. Let R_2 be the same segment with a topology defined by means of the following system of neighborhoods. All points $\neq 0$ get their ordinary neighborhoods whereas the neighborhoods of the point zero are the semisegments $[0, x]$ from which the points $x_n = 1/n$ are picked out. R_2 is an irregular Hausdorff space and the identical mapping of R_1 onto R_2 is a θ-homeomorphism, (the identical mapping of R_2 on R_1 is even continuous).

THEOREM 5. *A continuous image of a H-closed space is H-closed.*

The proof is easily given using the well known criterion of H-closedness due to Alexandroff: In order that the Hausdorff space R be closed, it is necessary and sufficient that from every covering of R by open sets $\{G\}$ a finite system G_1, $G_2, \cdots, G_n$ may be chosen in such a way that $\bar{G}_1 \cup \bar{G}_2 \cup \cdots \cup \bar{G}_n = R$.

THEOREM 6. *Let the two Hausdorff spaces X and Y have in common a point set R which is dense in X as well as in Y. Suppose that there exist θ-continuous mappings f and g of X on Y and of Y on X, respectively, which leave each point of R invariant. Then f and g are mutually inverse θ-homeomorphisms.*

PROOF. Let $y = f(x)$ and $g(y) = z$. Assume $x \neq z$. Choose the neighborhoods $U(x)$ and $U(z)$ without common points; then $U(x) \cap \overline{U(z)} = 0$. Take $U(y)$ such that $g(\overline{U(y)}) \subseteq \overline{U(z)}$ and $r \in R$ such that $r \in U(x), f(r) \in \overline{U(y)}$. Then

$$r = g(f(r)) \in U(x) \cap g\,(\overline{U(y)}) \subseteq U(x) \cap \overline{U(z)},$$

in contradiction with the choice of $U(x)$, $U(y)$. This contradiction proves the identity $x = z$ and thus the Theorem 6.

The proof of the following proposition presents no difficulty and may be left to the reader:

THEOREM 7. *If the space Y does not contain inseparable points (i.e. for any $y_1 \in Y$, $y_2 \in Y$ there are neighborhoods $U(y_1)$, $U(y_2)$ with $\overline{U(y_1)} \cap \overline{U(y_2)} = 0$) and Y is a θ-continuous image of X, then X also does not contain inseparable points.*

From the Theorems 5 and 7 follows that the H-closed spaces without inseparable points form a class invariant under θ-homeomorphisms. The above example shows that bicompactness is not invariant under θ-homeomorphisms.

5. The universal H-closed extension

We shall prove the following fundamental theorem:

THEOREM 8. *The space σR has the following properties:*

1°. *The space σR may be θ-continuously mapped on every H-closed extension of R in such a way that all points of R remain invariant under this mapping;*

2°. *Every H-closed extension of R possessing the property 1° is θ-homeomorphic with σR.*

PROOF. If 1° is proved, then 2° follows immediately by Theorem 6. Let us prove the assertion 1°. The construction of the mapping $f(x)$ below is the same as the corresponding construction in the paper of Alexandroff (proof of the Theorem II, pp. 408–409).

Let S be an arbitrary H-closed extension of R and

$$x = \{\Gamma_\alpha\} \in \sigma R.$$

Take the intersection

$$P_x = \prod_\alpha \bar{\Gamma}^S_\alpha$$

where $\bar{\Gamma}^S_\alpha$ means the closure of Γ_α in S.

We first prove that P is not vacuous. Denote by $G_\alpha = J(\bar{\Gamma}^S_\alpha)$ the set of all inner points of $\bar{\Gamma}^S_\alpha$. The system $\{G_\alpha\}$ is centered. Now two cases are possible.

a) The system $\{G_\alpha\}$ is not a Hausdorff system. This means that there exists a point $y_1 \in S$, such that for a certain $G_{\alpha_0} \in x$ the point y_1 does not belong to G_{α_0} while $y_1 \in \bar{G}^S_\alpha \subseteq \bar{\Gamma}^S_\alpha$ for all α, i.e. $y_1 \in P_x$.

b) The system $\{G_\alpha\}$ is a Hausdorff system. Then by Theorem 4 the G_α have a non-vacuous intersection obviously contained in P_x.

Now we prove that P_x consists of one point only. To this end we prove the

LEMMA 3. *If $y \in P_x$ then for each neighborhood $U(y)$*

$$R \cap U(y) \neq 0$$

PROOF. By the definition of the point y for every $\Gamma \in x$ we have $y \in \bar{\Gamma}^S$, wherefrom follows that for every $U(y)$ it is $(R \cap U(y) \cap \Gamma \neq 0$. The centered system $\{U(y)\}$ being a Hausdorff one, the same is true for the system $\{R \cap U(y)\}$. Consequently $\{\Gamma_\alpha\} \cup \{R \cap U(y)\}$ is a centered Hausdorff system. In virtue of the maximality of $\{\Gamma_\alpha\}$ it is $\{R \cap U(y)\} \subseteq \{\Gamma_\alpha\}$. The lemma is proved.

If now $y_1 \in P_x$, $y_2 \in P_x$ then choosing non intersecting neighborhoods $U(y_1)$ and $U(y_2)$ we see that $\{\Gamma_\alpha\}$ contains two non intersecting sets $R \cap U(y_1)$ and $R \cap U(y_2)$, which is impossible.

Putting for every $x = \{\Gamma_\alpha\} \in \sigma R$

$$y = f(x) = \prod \bar{\Gamma}_\alpha^S$$

we obtain a mapping of σR into S. To prove that f maps σR on the whole S take any $y \in S$. The system $\{U(y)\}$ is a centered Hausdorff system. Such is consequently the system $\{R \cap U(y)\}$ as well. Let us complete it to a Hausdorff end $\{\Gamma_\alpha\} = x$. Then

$$f(x) = \prod \bar{\Gamma}_\alpha^S \subseteq \prod \overline{(R \cap U(y))}^S \subseteq \prod \overline{U(y)}^S = y,$$

i.e. $f(x) = y$, which proves the assertion.

Let us finally prove that f is θ-continuous. Take $x \in \sigma R$,

$$f(x) = \prod \bar{\Gamma}_\alpha^S = y \in S$$

and an arbitrary neighborhood $U(y)$. We are looking for a neighborhood $U(x)$ such that $f(\overline{U(x)}) \subseteq U(y)$.

Let $G = U(y) \cap R$. As R is dense in S we have $\bar{G}^S = \overline{U(y)}^S$. By the Lemma 3 it is $G \in x$. Define

$$U(x) = U_G$$

To prove the relation $f(\overline{U(x)})^{\sigma R} \subseteq \overline{U(y)}$ take any $x' \in \overline{U(x)}^{\sigma R}$. Then

$$\Gamma' \cap G \neq 0 \text{ for an arbitrary } \Gamma' \in x'. \tag{1}$$

We have to prove that $y' = f(x') \in \overline{U(y)}^S$. Suppose it is not the case. Then there exists a neighborhood $U_1(y')$ such that $U(y) \cap U_1(y') = 0$. Take $\Gamma_0 = U_1(y') \cap R$. By the Lemma 3 it is $\Gamma_0 \in x'$. But $\Gamma_0 \cap G = 0$ in contradiction with (1).

Theorem 8 is thus completely proved.

6. Bicompact extensions

The present paragraph deals with bicompact extensions of the given space R. Therefore R must be supposed *completely regular*, since otherwise the space R does not possess Hausdorff bicompact extensions at all.

THEOREM 9. *Each of the spaces σR, αR, $\alpha' R$ may be continuously mapped on every bicompact extension B of R in such a way that all points of R remain invariant under this mapping.*

The proof of this theorem is quite similar to the proof of the first part of Theorem 8: it follows directly from the bicompactness of B that the intersection

$$P_x = \prod \bar{\Gamma}_\alpha^B$$

is non-vacuous for every $x = \{\Gamma_\alpha\}$. The mapping f on B will be continouus (and not only θ-continuous) by virtue of the regularity of B.

THEOREM 10. *If the space R is completely regular then $\alpha'_{\mathfrak{G}} R$ is also completely regular.*

PROOF. For the sake of simplicity we shall suppose that the basis $\mathfrak{G}$ is the system of all open sets of the space R, i.e. we consider the space $\alpha' R$.

We call the open set $\Gamma \subseteq R$ *canonic* if $\Gamma = R \backslash \overline{(R \backslash \bar{\Gamma})}$, i.e. if all inner points of $\bar{\Gamma}$ belong to Γ. Alexandroff has proved (3, p. 410) the following

LEMMA. *The sets U_Γ, where Γ is canonic, form a basis of $\alpha' R$.*

Let now $x = \{\Gamma_\alpha\}$ be an arbitrary point of the space $\alpha' R$. Let us show that if the set Γ_2 is completely regularly enclosed in a canonic open set Γ_1, then $\bar{U}_{\Gamma_2} \subseteq U_{\Gamma_1}$, i.e. $\bar{U}_{\Gamma_2} \cap (\alpha' R \backslash U_{\Gamma_1}) = 0$. Assume that this is not so. Then there is a point $\xi = \{G\}$ such that $\xi \in \bar{U}_{\Gamma_2} \cap (\alpha' R \backslash U_{\Gamma_1})$. The inclusion $\xi \in \bar{U}_{\Gamma_2}$ means that for each $G \in \xi$ the intersection $G \cap \Gamma_2$ is non-vacuous; $\xi \in U_{\Gamma_1}$ would mean that $\Gamma_1 \in \xi$, i.e. that there exists a $G \in \xi$ contained in Γ_1. Thus $\xi \bar{\in} U_{\Gamma_1}$ means that no $G \in \xi$ is contained in Γ_1, i.e. that none of the intersections $G \cap (R \backslash \Gamma_1)$ vanishes, and then (since Γ_1 is canonic) all $G \cap (R \backslash \Gamma_1)$ are different from zero. Thus, every $G \in \xi$ has a non vanishing intersection with Γ_2 as well as with $R \backslash \bar{\Gamma}_1$. In other words, *there exists a completely regular end, such that each of its coordinates has a non-vacuous intersection with both of the open sets Γ_2 and $R \backslash \bar{\Gamma}_1$, the closures of Γ_2 and $R \backslash \bar{\Gamma}_1$ being completely separated.* Let us lead this result *ad absurdum.*

LEMMA 4. *If G_1 and G_2 are open sets with completely separated closures and $G = (G_1 \cup G_2) \in \xi$, then either $G_1 \in \xi$, or $G_2 \in \xi$.*

PROOF. If the coordinate G' of the point ξ is contained in G, then G' falls into two sets G_1', and G_2', one of which is contained in G_1, and the other in G_2. Thus a completely regular centered system of sets $\xi = \{G'\}$ generates two systems $\{G_1'\}$ and $\{G_2'\}$. At least one of the systems $\xi \cup \{G_1'\}$ and $\xi \cup \{G_2'\}$ must be centered and completely regular, and must, consequently, coincide with ξ. The lemma 4 is proved.

LEMMA 5. *Let $f(x)$ be a function, equal to zero on $\bar{\Gamma}_2$ and to one on $R - \Gamma_1$; let $0 < a < b < 1$, and $\Gamma(a, b)$ be the open set, whose points x satisfy the condition $a < f(x) < b$; then every $G \in \xi$ has with $\Gamma(a, b)$ a non vacuous intersection.*[5] In fact, otherwise G would fall into two sets G_1 and G_2 with completely separated closures, such that $G_2 \cap \Gamma_2 = 0$ and $G_1 \cap (R - \bar{\Gamma}_1) = 0$, and, in virtue of Lemma 4,

[5] Hence, in particular, follows that $\Gamma(a, b)$ itself is non-vacuous.

either G_1, or G_2 would enter into ξ, and this contradicts the assumption that $\xi \in \bar{U}_{\Gamma_2} \cap (\alpha' R - U_{\Gamma_1})$. The Lemma 5 is proved.

Let us choose a_0 and b_0 such that $0 < a_0 < b_0 < 1$. Consider the aggregate of all $\Gamma(a, b)$, for which $0 < a < a_0 < b_0 < b < 1$. They form a completely regular centered system. The system $\xi \cup \{\Gamma(a, b)\}$ is, according to the above, centered, and since ξ and $\{\Gamma(a, b)\}$ are completely regular systems, $\xi \cup \{\Gamma(a, b)\}$ is also completely regular. But $\Gamma(a, b) \subseteq \Gamma_1$ and, consequently, $\Gamma(a, b) \cap (R - \bar{\Gamma}_1) = 0$ and $\xi \bar{\in} (\alpha' R - U_{\Gamma_1})$, i.e. $\bar{U}_{\Gamma_2} \subset U_{\Gamma_1}$. Thus we have proved that U_{Γ_2} is regularly enclosed in U_{Γ_1}. It is easily seen that this inclusion is also completely regular. In fact, because of the complete regularity of R, to every rational number t, $0 \leq t \leq 1$, may be put in correspondence an open set Γ_t in such a way that Γ_{t_1} is completely regularly enclosed in Γ_{t_2}, if $t_1 < t_2$, $\Gamma_t = \Gamma_1$ for $t = 1$ and $\Gamma_t = \Gamma_2$ for $t = 0$. By what has been proved, $\bar{U}_{\Gamma_{t_1}} \subseteq U_{\Gamma_{t_2}}$, if $t_1 < t_2$. Consequently U_{Γ_2} is completely regularly enclosed in U_{Γ_1}. We have proved that the space $\alpha' R$ is completely regular.

THEOREM 11. *If $\mathfrak{S}$ is an algebraically closed basis, then the space $\alpha'_{\mathfrak{S}} R$ is bicompact.*

The proof of this theorem is analogous to the proof of theorem 3, and will be therefore only sketched here. We use in this proof the following criterion for bicompactness of a completely regular topological space:

THEOREM 4′. *A completely regular space R is bicompact if and only if every completely regular centered system of elements of a given algebraically closed basis of R has a non-vanishing intersection.*

The necessity of this condition follows from Theorem 4.

To prove that this condition is sufficient, let us remember that by a known theorem of Tychonoff [1] a non-bicompact space R can be immersed in a completely regular space $R \cup \xi$ where ξ is a non-isolated point. Let $\{U(\xi)\}$ be the system of all neighborhoods of ξ, whose intersections with R belong to $\mathfrak{S}$:

$$(U(\xi) \cap R) \in \mathfrak{S}.$$

Then the system $\{U(\xi) \cap R\}$ is a completely regular centered system with a void intersection. Thus the announced criterion is proved.

We need furthermore the following

LEMMA 2′. *If $\mathfrak{S}'$ is an algebraically closed basis of R, then $\alpha'_{\mathfrak{S}} R - \bar{U}_G = U_{R-\bar{G}}$.*

This lemma is proved in the same way as Lemma 2. It follows from Lemma 2′:

COROLLARY. *The open sets U_G corresponding to an algebraically closed basis $\mathfrak{S}$ of the completely regular space R form an algebraically closed basis of $\alpha'_{\mathfrak{S}} R$.*

The rest of the proof of Theorem 11 is the same as that of Theorem 3.

From Theorems 9 and 11 follows immediately:

COROLLARY. *The space $\alpha' R$ is a maximal bicompact extension of R.*

Such an extension being unique,[6] the space $\alpha' R$ coincides with the Čech extension βR.

[6] See [3], Lemma IX.

REMARK. Thus we have given a *direct proof* of the fundamental identity $\alpha' R = \beta R$ for any completely regular R; the proof of this identity given in [3] makes essential use of other properties of the space βR.

Let us now see, under what conditions the space R has a *minimal* bicompact extension, i.e. an extension, which is a continuous image of every other bicompact extension of the space R. It is obvious that such an extension exists for every locally bicompact R, and is obtained by adjunction to R of one point. Let us prove that, conversely:

If the bicompact extension B of the space R is minimal, then it is obtained by adjunction to R of one point and, consequently, exists only for a locally bicompact R.

In fact, let B be a bicompact extension of the space R, such that $B - R$ contains more than one point. Let $x \in B - R$ and $y \in B - R$. Consider the space B', which is obtained from B by identifying the points x and y. This space is Hausdorff, bicompact and is an extension of R. The space B may be in an obvious manner continuously mapped on B' so that all points of R remain fixed. B cannot be a continuous image of B', since otherwise the mapping of B on B', indicated above, would be a homeomorphism (Theorem 6); but at the same time this mapping is not one-to-one.

Thus, our assertion is completely proved.

MATHEMATICAL INSTITUTE OF THE
ACADEMY OF SCIENCES OF THE U.S.S.R.

CITATIONS

[1] A. TYCHONOFF, *Über die topologische Erweiterung von Räumen*, Math. Annalen 102 (1929), 544.

[2] E. ČECH, *On bicompact spaces*, Annals of math. 38 (1937), 823–845.

[3] P. ALEXANDROFF, *Bikompakte Erweiterungen topologischer Räume*, Rec. Math. 5 (1939), 403–423.

ANNALS OF MATHEMATICS
Vol. 44, No. 3, July, 1943

FOUNDATIONS OF THE THEORY OF LIE GROUPS WITH REAL PARAMETERS

BY P. A. SMITH

In memory of E. R. VAN KAMPEN

(Received November 7, 1942)

The material presented below amounts roughly to the so-called fundamental theorems of Lie and their implications concerning Lie algebras, Lie subgroups and subalgebras. By straightforward and fairly elementary steps, we shall extend the concept of Lie group to include groups admitting coordinate systems in which the functions defining ab, the group product of elements a, b, possess continuous first order derivatives in the coordinates of a and satisfy a Lipschitz condition in those of b. That such groups are equivalent to classical (i.e. analytical) Lie groups was announced in 1936 by van Kampen, although apparently van Kampen published nothing by way of proof beyond a certain decisive uniqueness theorem concerning systems of ordinary differential equations. (This is contained in our theorem (12.1); the proof given below is essentially that of van Kampen [3]). It is conceivable that the condition relative to coordinates of b could be weakened or dispensed with entirely. For, the only part the condition in question plays in our development is to make it certain that the inverse of a certain mapping—the mapping into "canonical coordinates"—is single-valued.

In this connection mention should be made of the paper [1] of Garrett Birkhoff in which it is shown that the existence of continuous first order derivatives of ab with respect to the coordinates of both the a's and the b's is sufficient for defining a Lie group.

We shall consider only groups with a finite number of real parameters (whereas the groups considered by Birkhoff are not necessarily finite dimensional). This makes it possible to use the standard existence theorems for systems of ordinary real differential equations. We make no use whatever of the theory of partial differential equations.

We maintain throughout a purely local point of view in the sense that we consider only what happens in the neighborhood of the identity.

Historical notes. The proof of the uniqueness of square roots (8.5) is due to Claude Chevalley (unpublished).

The proof given below of Lie's theorem that a linear system of vector fields which is closed under commutation defines a Lie group is, we believe, due to van der Waerden [4]. We have supplied a number of preliminary lemmas which make the proof rigorously applicable to the case in which the vector fields are only assumed to possess continuous first order derivatives.

The theorem in §17 that every local subgroup of a Lie group is a Lie group is due essentially to E. Cartan [2].

That the functions ϕ_i defined in (14.16) are solutions of the so-called equations of Maurer (16.14) was first pointed out by Whitehead [5].

The existence of the functions a^t (a an element of a given group, t a real variable) was first established intrinsically by Garrett Birkhoff [1] although by methods quite different from those given below (8.6). A simple topological theorem which we require in this connection is proved in the appendix.

1. Local groups

The "groups" of classical Lie theory are not abstract topological groups in the contemporary sense since products are generally defined only in a neighborhood of the identity. Such partially defined groups have in recent years been called group nuclei, group germs, partial groups, local groups. Following Pontrjagin (*Topological Groups*) we adopt the name local group. It is to be understood that a *space* means a Hausdorff space.

DEFINITION. *A local group is a space $G = \{a, b, \cdots\}$ together with a function of composition, denoted by ab, satisfying the following conditions:* (1) *wherever defined, ab is single-valued and continuous in the pair (a, b); its values are elements of G;* (2) *if $a(bc)$, $(ab)c$ are defined, they are equal;* (3) *there exists a unique element* $\mathbf{1}$, *the identity, such that* $\mathbf{1}a$ *and* $a\mathbf{1}$, *when defined, are equal to* a; (4) *denoting by a^{-1} any element a' such that aa' and $a'a$ are defined and equal to* $\mathbf{1}$, *the function a^{-1} is single-valued and continuous wherever defined;* (5) *there exists a neighborhood U of* $\mathbf{1}$ *such that ab and a^{-1} are defined for every a, b in U. We shall call U a nucleus of G. Evidently every open subset of U which contains* $\mathbf{1}$ *is also a nucleus. G is a topological group if G itself is a nucleus.*

Let A, B be subsets of a local group G. Let $AB = \{ab, a \in A, b \in B\}$ assuming all these products ab are defined. Define A^{-1} similarly. It is an easy consequence of the relation $\mathbf{11} = \mathbf{1}$, the continuity of ab, a^{-1} and the properties of Hausdorff spaces that, given a nucleus V, there exists a nucleus U such that $\bar{U}\bar{U} \subset V$, $\bar{U}^{-1} \subset V$ (then of course $UU \subset V$, $U^{-1} \subset V$).—Every nucleus V contains a *symmetric* nucleus W ($W^{-1} = W$). In fact, take $W = V \cap V^{-1}$.

Let U, V be nuclei with $UU \subset V$. Using the associative law, we can define, as in the case of ordinary groups, products of all sets of three elements in U; evidently U can be so chosen that $UUU \subset V$. In fact, for given n, U can be so chosen that the n-fold product $U \cdots U$ is defined and in V.

If A, B are subsets of a local group and if A is open, it is easy to show that AB (if defined) is open.

Let V, V_1 be nuclei of local groups G, G_1. Let U be a nucleus of G such that $UU \subset V$, $U^{-1} \subset V$ and let U_1 be a similar nucleus in G_1. A continuous mapping $\tau: \tau U = U_1$ such that $\tau(ab) = \tau(a)\, \tau(b)$ is a *local homomorphism* $G \to G_1$. In particular τ is a local *isomorphism* $G \to G_1$ if the mapping $\tau U = U_1$ is a homeomorphism. It is obvious that a local homomorphism carries the identity into the identity and (in the neighborhood of the identity) inverses into inverses. The relation of local isomorphism is reflexive, symmetric, transitive.

Let G and H be local groups and assume that H is a *closed* subset of G and derives its topology from that of G. If the identities of H and G coincide and if the product of two elements of H, whenever defined, is the same as it is if the elements are regarded as elements of G, (provided this product is also defined) and if the same is true concerning inverses, we call H a *local subgroup* of G.

THEOREM (1.1). Let G, G_0 be local groups, f a mapping of a nucleus of G into a subset of G_0. Assume that there exists a nucleus V of G such that for a, b in V, $f(ab)$ and $f(a)f(b)$ are defined and equal, $f(a^{-1})$, $(f(a))^{-1}$ are defined and equal. Then there exists a nucleus U such that f is defined over $\bar{U}$ and such that under f, the images of sufficiently small nuclei of G are *open* in the set $\Gamma = f(\bar{U})$.

PROOF. Let X be a nucleus with $\bar{X}\bar{X} \subset V$, U a nucleus with $\bar{U}\bar{U} \subset X$. We shall show that if W is a nucleus such that $W \subset U$ and $W^{-1} \subset U$, then $f(W)$ is open in the set $\Gamma = f(\bar{U})$.

Let I consist of those elements i of $\bar{X}$ such that $f(i) = \mathbf{1}_0$. Evidently I is closed. Since $WI \subset U\bar{X} \subset X\bar{X} \subset V$, $f(WI)$ is defined. If $w \in W$, $i \in I$, then $f(wi) = f(w)f(i) = f(w)$ so that $f(WI) = f(W)$.

Consider an element a in $\bar{U}$ such that $f(a) \in f(WI)$. Then $f(a) = f(wj)$ say where $w \in W$, $j \in I$. Hence $f(a) = f(w)$, and hence $f(w^{-1}a) = (f(w))^{-1}f(a) = \mathbf{1}_0$. Now $w^{-1}a \in W^{-1}\bar{U} \subset U\bar{U} \subset \bar{X}$. Hence $w^{-1}a \in I$, say $w^{-1}a = i$. Then $a = wi \in WI$. In short, for elements a in $\bar{U}$, "$f(a) \in f(WI)$" implies "$a \in WI$", and accordingly, those elements not in WI are carried by f into elements not in $f(WI)$. Hence

$$f(\bar{U} - \bar{U} \cap WI) = f(\bar{U}) - f(WI) = \Gamma - f(W).$$

Now W being open, WI is open. Hence $\bar{U} - \bar{U} \cap WI$ is closed. Hence $f(W)$ is open in Γ.

(1.2) *Application.* Define the "product" of two elements of Γ in theorem (1.1) to be their product *as elements of* G_0 whenever this product exists and is an element of Γ. Similarly for inverses. Then we assert that Γ becomes a local subgroup of G_0 and f defines a local homomorphism $G \to \Gamma$. We have in fact only to show that Γ has a nucleus and that this nucleus is an image under f of a nucleus in G. Let W be a nucleus of G such that $WW \subset U$, $W^{-1} \subset U$. Let $x_0 = f(x)$, $y_0 = f(y)$ $(x, y \in W)$ be arbitrary elements of the open set $f(W)$. Then $f(x_0y_0) = f(xy) \in f(U) \subset \Gamma$ so that Γ-products are defined over $f(W)$. So are Γ-inverses.

For more details concerning local groups and subgroups see Pontrjagin, *Topological Groups.*

2. Coordinate systems

A local group G is *locally euclidean r-dimensional* if some nucleus is homeomorphic to euclidean r-space E_r.

A local euclidean local group G admits coordinate systems in the neighborhood of $\mathbf{1}$. The most convenient situation is of course that in which there are co-

ordinate systems covering the whole of G. For this reason we shall consider exclusively what we shall call *r-parameter local groups*, namely local groups which, as spaces, are homeomorphic to E_r. We denote such local groups generically by G_r. Aside from questions of convenience, justification for introducing the G_r's lies in the obvious fact that every locally euclidean r-dimensional local group is locally isomorphic to a G_r.

A *coordinate system* in G_r will always mean a coordinate system covering the *whole* of G_r,—that is, a coordinate system defined by a homeomorphism between E_r and the whole of G_r. It will always be understood that the origin of a coordinate system is at **1**. In a given coordinate system the i^{th} coordinate of an element will always be denoted by a superscript, as a^i, $(ab)^i$, $(a^2)^i$. The element **1** in a G_r with a coordinate system will always be denoted by O, the symbol for the vector $(0, \cdots, 0)$.

A coordinate system for G_r having been chosen, it will be convenient to regard the elements of G_r as vectors which can be added to each other and multiplied by scalars. We introduce also a modulus: $|a| = (\Sigma(a^i)^2)^{1/2}$. These vector functions depend of course on the given coordinate system and have no invariantive significance.

A function of one or more elements in G_r is generally understood to have values which are elements in G_r. In a given coordinate system, equations between functions and their derivatives stand for a set of equations corresponding to the different coordinates. Thus, $f(a, b) = c$ means $f^i(a, b) = c^i$; $\partial(ab)/\partial a^j = c$ means $\partial(ab)^i/\partial a^j = c^i$ $(i = 1, \cdots, r)$.

We shall presently be dealing with single-valued functions $f(a, b, \cdots c)$ where some of the variables are real variables, some are elements in a G_r with definitely chosen coordinate system, and where the values of f are real or in G_r. To avoid unessential details we shall not always specify the domain of existence of such a function. But in any case this domain will always contain a domain $D_\delta : |a| < \delta, \cdots, |c| < \delta$. We shall say that f is *locally continuous* if it is continuous in the totality of its variables over some D_δ. The same definition is to apply to functions $f(a, \cdots, c; e)$ when e is an arbitrary *unit vector* in G_r, except that the relation $|e| = 1$ is to be adjoined to the definitions of D_δ.

A coordinate system Σ for G_r is *left-differentiable* if the functions

$$(ab)^i = f^i(a^1, \cdots, a^n; b^1, \cdots, b^n)$$

possesses locally continuous derivatives with respect to $a^1, \cdots, a^r$. Σ is of *class* $k(k \geqq 1)$ if the f^i possess locally continuous derivatives of orders less than or equal to k with respect to the a's and b's. Σ is *analytic* if there is a $\delta > 0$ such that the f^i can be represented as power series which converge for $|a| < \delta$, $|b| < \delta$.

3. Left differentiable coordinates

Consider a G_r with a left-differentiable coordinate system. Let e be an arbitrary unit vector, s a real variable, and let d/ds_0 denote the s-derivative at $s = 0$.

Let

$$\psi_i(b) = \left(\frac{\partial(ab)}{\partial a^i}\right)_{a=O} \tag{3.1}$$

$$\psi(b; e) = \frac{d}{ds_0}((se)b) = \Sigma e^i \psi_i(b) \tag{3.2}$$

$$\psi(s, b; e) = \frac{d}{ds}((se)b) = \Sigma e^i \left(\frac{\partial(ab)}{\partial a^i}\right)_{a=se}. \tag{3.3}$$

Left differentiability implies that these three functions are locally continuous. Evidently

$$\psi(0, b; e) = \psi(b; e) = \lim_{s \to 0} \frac{1}{s}((se)b - b). \tag{3.4}$$

Putting $b = O$ gives

$$\psi(O; e) = e. \tag{3.5}$$

We note also that

$$\psi_j^i(O) = \left(\frac{\partial(aO)^i}{\partial a^j}\right)_{a=0} = \frac{\partial a^i}{\partial a^j} = \delta_j^i . \tag{3.6}$$

(3.7) *In a left differentiable coordinate system*

$$ab = a + b + |a| F(a, b)$$

where F is locally continuous and $F(O, O) = O$.

PROOF. Let

$$\begin{aligned} &\Omega(t, s, b; e) = \frac{1}{t}[((s + t)e)b - (se)b] - \psi(s, b; e), \qquad t \neq 0; \\ &\Omega(0, s, b; e) = 0. \end{aligned} \tag{3.8}$$

Ω is defined over some domain $D_\delta : |s|, |t|, |b| < \delta, |e| = 1$, and is continuous, evidently, at the places where $t \neq 0$. We assert that Ω is locally continuous. It is sufficient to show that as $t \to 0$, Ω converges to O *uniformly* with respect to s, b, e when $|s| < \delta, |b| < \delta, |e| = 1$. We have

$$\frac{1}{t}[((s + t)e) - (se)b]^i = \frac{1}{t}\int_s^{s+t} \psi^i(\sigma, b; e)\, d\sigma = \psi(\mu^i, b; e)$$

where $s \leqq \mu^i \leqq s + t$. Since $\psi(s, b; e)$ is locally continuous, $(1/t)[\quad]^i$ is uniformly near $\psi(s, b; e)$ when $|t|$ is small, which proves our assertion. Now put $s = 0$ in (3.8), multiply by t and use (3.4). We obtain

$$(te)b = b + t\psi(b; e) + t\Omega(t, 0, b; e). \tag{3.9}$$

By (3.5) and the local continuity of $\psi(b; e)$, we may write $\psi(b; e) = e + E(b; e)$

where E is locally continuous and $E(O; e) = O$. On substituting into (3.9) we obtain

$$(3.10)\qquad \begin{aligned}(te)b &= te + b + tE(b; e) + t\Omega(t, 0, b; e)\\ &= te + b + tE'(t, b; e)\end{aligned}$$

where E' is locally continuous and $E'(0, O; e) = O$. On putting $a = te$ and allowing only non-negative values of t, so that $|a| = t$, (3.10) becomes $ab = a + b + |a| F(a, b; e)$ where F is locally continuous and $F(O, O; e) = O$. Evidently e cannot enter the function F effectively; hence it can be omitted. This completes the proof of (3.7).

From (3.7) we obtain: $a^2 = 2a + (a)F(a, a)$, $a^3 = aa^2 = a + a^2 + |a| F(a, a^2) = 3a + |a| (F(a, a) + F(a, a^2))$, and so on. Hence

$$(3.11)\qquad a^n = na + |a| (F(a, a) + F(a, a^2) + \cdots + F(a, a^{n-1})).$$

All the terms in this formula are well defined if a is restricted to a sufficiently small nucleus. In case $n = 2$, we may write simply

$$(3.12)\qquad a^2 = 2a + |a| F(a),$$

where $F(a)$ is locally continuous and $F(O) = O$

Putting $b = a^{-1}$ in (3.7), we obtain

$$(3.13)\qquad a^{-1} = -a - |a| F(a, a^{-1}).$$

4. Analytic coordinates

The constant terms in the power series expansions for the functions $(ab)^j$ in an analytic coordinate system must vanish since $(O, O)^j = 0$. To compute the linear terms, we note that $(\partial ab)^i/\partial a^j)_{a=O,\, b=O} = \delta_j^i$ (see (3.2)). By symmetry, the same relations hold if a^i is replaced by b^i. Hence we have

THEOREM (4.1). *In an analytic coordinate system the power series which define ab are of the form*

$$(ab)^j = a^j + b^j + \cdots \qquad (j = 1, \cdots, r)$$

where the dots stand for terms of degree greater than 1.

THEOREM (4.2). *Relative to an analytic coordinate system for G_r let $\mathfrak{G}_h, \mathfrak{G}_k, \cdots, \mathfrak{G}_m$ $(h + k + \cdots + m = r)$ be linear subspaces which contain O and together span G_r. There exists a nucleus V such that if $a, b, \cdots c$ are arbitrary elements in $V \cap \mathfrak{G}_h$, $V \cap \mathfrak{G}_k$, $\cdots V \cap \mathfrak{G}_m$, the correspondence*

$$a + b + \cdots + c \rightarrow ab \cdots c$$

is a homeomorphism. Every element near O is therefore uniquely expressible in the form $ab \cdots c$.

Suppose for example that we have two spaces, say $\mathfrak{G}_h$, $\mathfrak{G}_k$ with $h + k = r$. We may perfectly well assume that $\mathfrak{G}_h$ is the $(a_1, \cdots, a_h)$-coordinate space,

$\mathfrak{G}_k$ its linear complement. Then $a' = (a^1, \cdots, a^k, 0, \cdots, 0)$, $a'' = (0, \cdots, 0, a^{k+1}, \cdots, a^r)$ are arbitrary elements of $\mathfrak{G}_k$, $\mathfrak{G}_k$, the product $c = a'a''$ is a function of $(a^1, \cdots, a^r) = a' + a''$. It is sufficient to show that the jacobian matrix $J(a) = (\partial c^i/\partial a^j)$ is non-singular at $a = O$. With the aid of (4.1) we find immediately that $J(O)$ is the identity matrix.

5. Canonical coordinates

Let t, s be real variables. A coordinate system Σ for G_r is *canonical* if the relations

$$(5.1) \qquad (ta)(sa) = (t + s)a; \qquad \frac{d}{dt_0}((ta)(tb)) = a + b$$

hold in some neighborhood of $a = O$, $b = O$, $t = 0$. Here, as always, d/dt_0 means the derivative at $t = 0$.

We note that $(5.1)_2$ is always implied by left-differentiability. For, by (3.7),

$$(5.2) \qquad \frac{1}{t}((ta)(tb)) = a + b \pm |a| F(ta, tb) \to a + b$$

as $t \to 0$. Hence

(5.3) *A left differentiable coordinate system in which* $(5.1)_1$ *holds is canonical.*

It is evident that if two coordinate systems for G_r are linearly related to each other and if one is canonical, so is the other. A sort of converse is

THEOREM (5.4). *A local homomorphism* τ: $G_r \to G'_{r'}$, *expressed in terms of canonical coordinates for* G_r *and* $G'_{r'}$, *is linear over some nucleus of* G_r.

PROOF. The theorem asserts that there exists a nucleus V of G_r such that

$$(5.5) \qquad \tau(ta) = t\tau(a), \qquad \tau(a + b) = \tau(a) + \tau(b)$$

whenever a, b, ta are in V. To simplify the exposition we shall make no further mention of the domains in which functions and operations are defined. (The missing details are not far from trivial.)—Since coordinates are canonical.

$$(5.6) \qquad \tau(nb) = \tau(b^n) = (\tau(b))^n = n\tau(b)$$

where n is a positive or negative integer. Writing $a = nb$, we obtain $\tau(a/n) = \tau(a)/n$ where a/n means $(1/n)a$. If we now apply (5.6) with n replaced by m and b by a/n we get

$$\tau\left(\frac{m}{n}a\right) = \frac{m}{n}\tau(a).$$

Hence $(5.5)_1$ follows by continuity. Suppose that under the correspondence τ, $a \to a_0$ and $b \to b_0$. By what we have just proved, $ta \to ta_0$, $tb \to tb_0$. Hence $(ta)(tb) \to (ta_0)(tb_0)$ and hence $t^{-1}((ta)(tb)) \to t^{-1}((ta_0)(tb_0))$. On letting t approach 0, we have, using $(5.1)_2$, $a + b \to a_0 + b_0$ which is $(5.5)_2$.

THEOREM (5.7). *Let* G_r *be a local group with an analytic canonical coordinate*

system. Let H be a local subgroup of G_r. There exists in G_r a linear subspace $\mathfrak{G}$, possibly 0-dimensional, and a nucleus W such that $\mathfrak{G} \cap W = H \cap W$. Evidently $\mathfrak{G}$ may be regarded as a local subgroup of G_r, locally identical with H. Any coordinate system for $\mathfrak{G}$ which is linearly related to that of G_r is an analytic canonical system for $\mathfrak{G}$.

PROOF. The second part of the theorem is evident once the existence of $\mathfrak{G}$ and W is established.

Let V be a spherical nucleus of G_r so small that $V \cap H$ is a nucleus of H. Let λ be a line through O such that the segment $\lambda \cap V$ is contained in H. Let $\mathfrak{G}$ be the point-set union of the lines λ if any exist, otherwise let $\mathfrak{G} = \{O\}$. We assert that $\mathfrak{G}$ is a linear subspace of G_r. It will be sufficient to show that when $\mathfrak{G}$ does not merely consist of O, the sum of two elements of $\mathfrak{G}$ near O is an element of $\mathfrak{G}$. Let U be a nucleus so small that products and vector sums of pairs of elements in U are in V. Let a, b be elements in $\mathfrak{G} \cap U$. Then $a + b \in \mathfrak{G} \cap V$. Since a, b are also in $H \cap U$, we have $(a/n)(b/n) \in H \cap V$ where a/n means $(1/n)a$. From $(5.1)_2$, $n((a/n)(b/n))$ is arbitrarily near $a + b$ if n is large enough. Hence for large n the elements $m((a/n)(b/n))$, $m = -n, -n + 1, \cdots, n$, lie close together along a linear segment beginning near $-(a + b)$ and ending near $a + b$. Hence the segment σ joining $-(a + b)$ to $a + b$ consists of cluster points of points $m((a/n)(b/n))$. Coordinates being canonical, these points are mth powers of elements $(a/n)(b/n)$ which are in H, hence they themselves, as well as the points of σ, are in H. Hence $a + b \in \mathfrak{G}$.

We can now prove that $\mathfrak{G} \cap W = H \cap W$ for some nucleus W. Suppose no such W exists. Then there exists a sequence $\{a_i\}$ converging to O, where a_i is in H but not in $\mathfrak{G}$. Let $\mathfrak{G}'$ be the linear complement of $\mathfrak{G}$ and write $a_i = g_i g_i'$ ($g \in \mathfrak{G}, g' \in \mathfrak{G}'$) in accordance with (4.2). (In case $\mathfrak{G} = \{O\}$, take $g_i = O, g_i' = a_i$). Since $g_i' \to O$, the positive and negative multiples of g_i', i large, lie close together along a line through O. Hence there is a line μ through O consisting of cluster points of the set $\{mg_i'\}$ ($m = \pm 1, \pm 2, \cdots, i = 1, 2, \cdots$). Since these points are in $\mathfrak{G}'$, so is μ. We shall show however that μ is in $\mathfrak{G}$, a contradiction which completes the proof of the theorem. Let p be a point of $\mu \cap V$. We may write $p = \lim m_j g_j'$ as $j \to \infty$, where $m_j \to \infty$ and $\{g_j'\}$ is a subsequence of $\{g_i'\}$. Since $a_j \in H$ and since $g_j \to O$ so that $g_j \in H$ when j is large, elements $g_j' = g_j^{-1}a_j$ with large j are in H. Hence for large j, $m_j g_j' = (g_j')^{m_j} \in H$. Hence $p \in H \cap V \subset \mathfrak{G}$.

6. The full matric group

We recall the following elementary facts about real square matrices. Let $X = (X^{ij})$ be such a matrix. The series

$$e^{tX} = I + tX + (t^2/2!)X^2 + \cdots,$$

where I is the identity matrix and t a real variable, converges for every X and t. Evidently the elements $(e^{tX})^{ij}$ are power series in the variables X^{ij}, t, con-

vergent for all values of these variables. It can be seen by direct substitution that $Y = e^{tX}$ satisfies the equation

$$\frac{dY}{dt} = XY;$$

Y is, in fact, that (unique) solution which satisfies the initial condition $Y(0) = I$. Since $Y(t + s)$ and $Y(t)Y(s)$ are solutions which equal $Y(s)$ when $t = 0$, we have: $e^{tX}e^{sX} = e^{(t+s)X}$.

Now let $M^r = \{I, A, \cdots\}$ be the group of real r-rowed non-singular square matrices. An element A of M^r may be regarded as a point in E_{r^2} with coordinates $A^{11}, A^{12}, \cdots, A^{rr}$ written in some definite order. M^r is thus an open subset of E_{r^2} and may itself be regarded as a space, subspace of E_{r^2}. Since the functions AB, A^{-1} are continuous, M^r is a topological group. We call $A^{11}, \cdots, A^{rr}$ the *natural coordinates* of A in M^r.

THEOREM (6.1). *There exists an r^2-parameter local group M_{r^2} which has an analytical canonical coordinate system Σ, and is locally isomorphic to M^r. In terms of Σ and the natural coordinates of M^r, there is a local isomorphism $\tau: M^r \to M_{r^2}$ which is analytic in the neighborhood of the identity.*

PROOF. Let $E_{(ij)}$ be the r-rowed matrix which contains a 1 in the (i, j)-position and zeros elsewhere. Let $X(x) = e^E$ where $E = \Sigma x^{ij}E_{(ij)}$. The natural coordinates $X^{pq}(x)$ are everywhere convergent power series in $x^{11}, \cdots, x^{rr}$. Now let $x^{11}, \cdots, x^{rr}$ be regarded as coordinates in a euclidean space M_{r^2}. The analytic correspondence $\sigma: x \to X(x)$ maps a neighborhood of the origin of M_{r^2} into a point set in M^r containing I. This mapping is in fact a homeomorphism. To show this it is sufficient to show that the jacobian matrix of σ is non-singular at $x = 0$. We have

$$\left(\frac{\partial X^{ij}(x)}{\partial x^{pq}}\right)_{x=0} = \frac{d}{dx^{pq}}\left(e^{x^{pq}E_{(pq)}}\right)^{ij}_{x^{pq}=0} = (E_{(pq)})^{ij}$$

which equals 1, when $(ij) = (pq)$, zero otherwise. The matrix in question is therefore the r^2-rowed identity at $x = 0$.—Let $\tau: X(x) \to x$ be the inverse of σ. With the aid of τ we carry group products of elements in a suitable nucleus of M^r isomorphically over into M_{r^2} converting M_{r^2} into a local r^2-parameter group. The coordinate system Σ: $x^{11}, \cdots, x^{rr}$ in M_{r^2} has the required properties. Analyticity of Σ follows from the analyticity of τ and the fact that products in M^r are defined by polynomials in the natural coordinates. That E is canonical follows from (5.3) and the identity $e^{tX}e^{sX} = e^{(t+s)X}$.

7. The function $a^{-1}ba$

This function of a and b plays an important part in the analytical theory of local groups.

THEOREM (7.1). *In canonical coordinates the function $a^{-1}ba$ is linear in b and analytic in a. More precisely, there exists a $\delta > 0$ such that for $|a| < \delta$,*

$|b| < \delta$, the coordinates $(a^{-1}ba)^i$ are linear functions of $b^1, \cdots, b^r$ with coefficients which are power series in $a^1, \cdots, a^r$ convergent for $|a| < \delta$.

PROOF. Let $f(t) = a^{-1}(tb)a$ where a, b are fixed elements near O, t a real variable. The points $f(t)$ with $-1 \leqq t \leqq 1$ constitute a simple arc γ through O. Since coordinates in G_r are canonical $f(s)f(t) = f(s + t)$, $(f(t))^{-1} = f(-t)$, for small t, s. Hence γ is a local subgroup of G_r. By (5.7), γ may be regarded as imbedded in a local subgroup G which is a line through O. We may write $G = \{se\}$ where e is a unit vector in G. The real parameter s is a canonical coordinate for G. Now let t be regarded as a canonical coordinate for R, the one-parameter additive group of real numbers. Evidently $f(t)$ defines a local isomorphism $R \to G$. In terms of s and t, this isomorphism is given by $s = kt$, $k \neq 0$ (5.4). Hence $f(t) = kte$. In particular, $f(1) = ke$ so that $f(t) = tf(1)$, or $a(tb)a^{-1} = t(a^{-1}ba)$, proving that $a^{-1}ba$ is homogeneous in b.

We next show that $a^{-1}ba$ is additive in b. Let $b' = a^{-1}ba$, $c' = a^{-1}ca$. Using $(5.1)_2$ and the homogeneity just established we have, as $t \to 0$,

$$\begin{aligned} b' + c' &= \lim \frac{1}{t}\,((tb')(tc')) = \lim \frac{1}{t}\,((a^{-1}(tb)a)(a^{-1}(tc)a)) \\ &= \lim \frac{1}{t}\,(a^{-1}(tb)(tc)a) = \lim a^{-1}\left(\frac{1}{t}\,((tb)(tc))a\right) \\ &= a^{-1}(b + c)a. \end{aligned}$$

We have now shown that for given a, $a^{-1}ba$ is linear in b. Let $A(a)$ be the matrix of coefficients of the transformation $b \to a^{-1}ba$. We show that the elements of $A(a)$ are convergent power series in $a^1, \cdots, a^r$. Since $A(ac) = A(a)A(c)$, the matrices $A(a)$, with $|a|$ small, are elements of the full matric group M^r. Consider the local group M_{r^2} with canonical coordinates $x^{11}, \cdots, x^{rr}$ and the local analytic isomorphism $\tau: M^r \to M_{r^2}$ of theorem (6.1). Let $\sigma(a) = \tau(A(a))$. The set $\Gamma = \{\sigma(a), |a| \leqq \delta\}$ is defined and closed (in M_{r^2}) if δ is sufficiently small. Since $\sigma(ac) = \sigma(a)\sigma(c)$ and $\sigma(a^{-1}) = \sigma(a))^{-1}$, Γ is a local subgroup of M_{r^2} according to (1.1) and (1.2), and σ defines a local homomorphism, $G_r \to \Gamma$. By theorem (5.7), Γ coincides, in the neighborhood of its identity, with a linear subspace N of M_{r^2}. Hence Γ may be considered as imbedded in N, hence admits canonical coordinates $y^1, \cdots, y^r$ linearly related to $x^{11}, \cdots, x^{rr}$ (5.7), say $x^{ij} = L^{ij}(y)$. Now τ is given say by

$$A^{ij} = \psi^{ij}(x^{ii}, \cdots, x^{rr})$$

where the ψ^{ij} are convergent power series and the A^{ij} are natural coordinates in M^r. Let $x^{ij}(a)$ be the x^{ij}-coordinates of $\sigma(a)$, $y^i(a)$ the y^i-coordinates of $\sigma(a)$. Then $x^{ij}(a) = L^{ij}(y^1(a), \cdots, y^r(a))$ and

$$A^{ij}(a) = \psi^{ij}(L^{11}(y^1(a), \cdots, y^r(a)), \cdots).$$

Now in terms of $a^1, \cdots, a^r$ and $y^1, \cdots, y^r$, σ is linear (5.4). Hence $y^i(a)$ is linear in $a^1, \cdots, a^r$ and hence the $A^{ij}(a)$ are convergent power series in $a^1, \cdots, a^r$. This completes the proof.

We obtain an explicit expansion for $a^{-1}ba$. Denote by $\mathfrak{A}_t$ the linear transformation $b \to (ta)^{-1}b(ta)$. Evidently $\mathfrak{A}_t\mathfrak{A}_s = \mathfrak{A}_{s+t}$. Let

$$(7.2) \qquad [a, b] = \frac{d}{ds_0} \mathfrak{A}_s b = \frac{d}{ds_0} (sa)^{-1} b(sa).$$

Evidently $[a, b]$ is linear in b. Let A be the linear transformation $b \to [a, b]$. We have

$$\frac{dA_t b}{dt} = \lim_{s \to 0} (\mathfrak{A}_{s+t} - \mathfrak{A}_t)b = \lim_{s \to 0} \frac{1}{s} (\mathfrak{A}_s - I)\, \mathfrak{A}_t b = A\mathfrak{A}_t b.$$

Hence (see 5.6) $\mathfrak{A}_t = e^{tA}$. Writing $\mathfrak{A}_1 = \mathfrak{A}$, we have

$$(7.3) \qquad a^{-1}ba = \mathfrak{A}b = e^A b = b + [ab] + \frac{1}{2!}[a[ab]] + \cdots .$$

The function $[a, b]$, linear in b, is called the *commutator* of a, b. We shall see later that $[a, b]$ is also linear in a.

8. Existence of the 1-parameter subgroups $a(t)$

The propositions in this section concern a G_r with left-differentiable coordinate system. V being a nucleus of G_r let

$$F(V) = \text{l.u.b. } \{F(a, b),\ a \in V,\ b \in V\}$$

where F is defined in (3.7). Local continuity of F and the relation $F(O, O) = O$ imply that $F(V) \to 0$ as diameter $V \to 0$.

(8.1) Let V_δ be a spherical nucleus of radius δ such that $F(V_\delta) < 1$. Let $a_1, a_2, \cdots$ be a sequence of elements in V_δ such that $|a_n| < (\delta/8)(3/4)^n$. Then the products

$$a_1,\ a_2a_1,\ a_3(a_2a_1),\ a_4(a_3(a_2a_1)),\ \cdots$$

(call them $p_1, p_2, \cdots$) are defined and

$$|p_n| < \frac{\delta}{4}\left(1 + \frac{3}{4} + \cdots \left(\frac{3}{4}\right)^n\right).$$

Hence in particular $p_n \in V_\delta$.

PROOF. Assume that p_{n-1} is defined and satisfies the inequality. Then $p_n = a_np_{n-1}$ is defined and by (3.7)

$$p_n = a_n + p_{n-1} + |a_n| F(a_n, p_{n-1}).$$

Since $|F(a_n, b_{n-1})| < 1$, $|p_n - p_{n-1}| < 2|a_n| < (\delta/4)(3/4)^n$. Hence

$$|p_n| \leqq |p_n - p_{n-1}| + |p_n| < \frac{\delta}{4}\left(\frac{3}{4}\right)^n + \frac{\delta}{4}\left(1 + \frac{3}{4} + \cdots + \left(\frac{3}{4}\right)^{n-1}\right)$$
$$= \frac{\delta}{4}\left(1 + \cdots \left(\frac{3}{4}\right)^n\right).$$

Let Z denote a definitely chosen nucleus such that $F(Z) < 1/4$.

(8.2) If $O \neq a \in Z$, then (i) $|a| \leqq 3/4 \, |a^2|$ and (ii) the angle θ between the vectors $\overrightarrow{aa^2}$ and $\overrightarrow{a, 2a}$ is less than a right angle.

Proof. Using (3.12) and the fact that $F(Z) < 1/4 < 2/3$ we have

$$|2a| \leqq |a^2| + \tfrac{2}{3}|a|$$

when $a \in Z$. Transposing the last term and multiplying by 3/4 gives (i). As for (ii), we have $|a^2 - 2a| < |a|$ (3.12). Hence in the triangle formed by the vectors in question, the side opposite θ is shorter than one side adjacent to θ. Hence (ii) is true.

(8.3) Not every power of an element in Z is in Z. For if $a \in Z$ and $n \to \infty$,

$$|a^{2^n}| \geqq |\tfrac{4}{3}|^n \, |a| \to \infty.$$

(8.4) Every element a in Z has a square root $a^{1/2}$ such that $|a^{1/2}| \leqq (3/4)\,|a|$.

Proof. Let σ be the mapping $a \to a^2$. If b is a point on the spherical boundary of Z, then $b \neq \sigma b$ and the vector $\overrightarrow{b, \sigma b}$ makes an acute angle with the outward normal vector at b (8.2) from which it follows (see appendix) that $Z \subset \sigma(Z)$. Hence every element a in Z is the square of at least one element $a^{1/2}$ in Z. The stated inequality for $a^{1/2}$ follows from (8.2).

We shall call $a^{1/2}$ a *principal* square root of a if $|a^{1/2}| \leqq |a|$. Proposition (8.4) implies that every element in Z has a principal square root in Z.

(8.5) Z contains a nucleus W the principal square roots of whose elements are unique.

Proof. Let Z', Z'' be symmetric nuclei such that $\bar{Z}'' \subset Z'$, $\bar{Z}' \subset Z$. Let

$$M = \text{g.l.b.} \{|a - a^{-1}|, \quad a \in Z' - Z''\}.$$

Evidently $M > 0$, otherwise Z would contain an element of order 2 which (8.2) shows to be impossible. Now choose a symmetric nucleus $Y \subset Z''$ such that $|z - yzy^{-1}| < M$ when $z \in Z$, $y \in Y$. The successive powers of an element y in Y lie more closely together as y is nearer O. Hence if Y is small enough, the first element in the sequence $y, y^2, \cdots$ which is not in Z''—there will be such an element by (8.3)—must lie in $Z' - Z''$. Assume Y to be so chosen. Let W be a spherical nucleus $W \subset Y$. We assert that the principal square roots of elements in W are unique. If not, there exist elements a, b, c in W such that $a^2 = b^2 = c$, $a \neq b$. Let $h = ab^{-1}$. Since $h \in Y$, the smallest power h^n of h which fails to be in Z'' will be in $Z' - Z''$. Now $a = hb$ so that $b^2 = a^2 = hbhb$, $h = b^{-1}h^{-1}b$, $h^n = b^{-1}h^{-n}b$. Since Z', Z'' are symmetric, $h^{-n} \in Z' - Z''$. Hence $|h^n - h^{-n}| > M$. But since $b \in Y$, $|h^{-n} - b^{-1}h^{-n}b| < M$, a contradiction.

In the following proposition V_ϵ denotes a spherical nucleus of radius ϵ.

(8.6) Let δ be a number not exceeding the radius of W. For every element a in $V_{\delta/8}$ there is a uniquely determined function $a(t)$ whose values are elements

in V_δ, defined for $0 \leqq t \leqq 1$ with $a(0) = O$, $a(1) = a$, and such that $a(t)a(s) = a(s + t)$. If $a \neq O$, the arc defined by $a(t)$ is simple.

PROOF. Let a be an element in $V_{\delta/8}$ and let $a^{1/2}$ be the unique principal square root of a. Since $a^{1/2}$ is also in $V_{\delta/8}$, it has a unique principal square root $a^{1/4}$ in $V_{\delta/8}$, and so on.

Let m, n be non-negative integers with $m < 2^n$. We may write

$$\frac{m}{2^n} = \frac{x_1}{2} + \frac{x_2}{2^2} + \cdots + \frac{x_n}{2^n}, \qquad x_i = 0 \text{ or } 1.$$

Since $|a| < \delta/4$ and $F(V_{\delta/8}) < F(Z) < 1/4$, we have (8.4)

$$|a^{x_i/2^n}| < \left(\frac{3}{4}\right)^n \frac{\delta}{8}.$$

Writing $x = x_1/2$, $y = x_2/4$, $\cdots$ it follows from (8.1) that

$$(a^{1/2^n})^m = a^z(\cdots(a^y a^x)\cdots)$$

is defined and in V_δ. The preceding statements evidently hold if δ is replaced by a smaller number. Hence we have proved the following proposition:

A) If a is an element in $V_{\epsilon/8}$ where $\epsilon \leqq \delta$ then $(a^{1/2^n})^m$ is defined and in V_ϵ.

Now let P be the set of rational numbers $\{m/2^n, m < 2^n\}$ dense on the interval (0, 1). We see readily that for given a, the element $(a^{1/2^n})^m$ depends only on the value of $m/2^n$. Hence we write $(a^{1/2^n})^m = a^{m/2^n}$. The function $\alpha(p) = a^p$ is single-valued over P. Evidently $\alpha(p)\alpha(q) = \alpha(p + q)$.

We assert that $\alpha(p) \to O$ as $p \to 0$. Take $\epsilon > 0$ and n so large that $|\alpha(1/2^n)| < \epsilon/8$. Now the values of p which are larger than $1/2^n$ are of the form $(1/2^n)(h/2^k)$, $0 \leqq h \leqq 2^k - 1$. Hence we may apply proposition A) with a replaced by $\alpha(1/2^n)$. We conclude that $|\alpha(p)| \leqq \epsilon$ when $p < 1/2^n$. Hence $\alpha(p) \to O$.

We assert next that $\alpha(p)$ is uniformly continuous over P. For take $\epsilon > 0$. From the preceding paragraph and group continuity there exists a number μ independent of p such that $|\alpha(p)\alpha(q) - \alpha(p)| < \epsilon$ when $q < \mu$. Hence $|\alpha(p + q) - \alpha(p)| < \epsilon$ when $q < \mu$, proving our assertion.—From this it follows, by elementary continuity considerations, that there exists a unique continuous single-valued function $a(t)(0 \leqq t \leqq 1)$ such that $a(p) = \alpha(p)$ over P. Evidently the relation $a(p)a(q) = a(p + q)$ extends itself to the whole interval (0, 1) by continuity. Moreover, $a(t) \in \bar{V}_\delta$ since $\alpha(p) \in V_\delta$.

It remains only to be shown that $a(t) \neq a(s)$ when $t \neq s$, $a \neq O$. Suppose on the contrary that $a(t + h) = a(t)$, $h \neq 0$. Then $a(t)a(h) = a(t)$ so that $a(h) = O$. From the relation $|a| \leqq 3/4\,|a^2|$ we have $|a(h/2)| \leqq |(a(h/2))^2| = |a(h)| = 0$. Hence $a(h/2) = O$ and similarly $a(h/2^n) = O$. Taking n large there exists an m such that $mh/2^n$ is near 1, hence such that $(a(h/2^n))^m$ is near a. We conclude that $a = O$, a contradiction.

Let U be the spherical nucleus whose radius is one eighth the radius of W. The function $a(t)$ is defined for $a \in U$, and the elements $a(t)$ are in $\bar{W}$.

(8.7) There exists a positive number T depending on a ($a \in U$) such that

$$\frac{1}{2}|a| \leqq \left|\frac{1}{t}a(t)\right| \leqq 2|a|$$

when $0 < t \leqq T$.

The proof of the first inequality is essentially the same as that of the second. Suppose the second is false, that is, suppose that $|a(t_i)/t_i| > 2|a|$ for some a and some sequence $t_i \to 0$. For each i choose an integer n_i such that

$$\frac{1}{n_i + 1} \leqq t_i \leqq \frac{1}{n_i}.$$

Then $n_i t_i \to 1$ as $i \to \infty$. If we refer to (3.7) and use the fact that $F(U) < 1/4$ (since $U \subset Z$) we have

$$\begin{aligned} a(n_i t_i) &= (a(t_i))^{n_i} \\ (8.8) \qquad &= n_i a(t_i) + |a(t_i)|[F(a(t_i), a(t_i)) + \cdots + F(a(t_i), a((n_i - 1)t_i))] \\ &= n_i a(t_i) + n_i |a(t_i)| H, \text{ say,} \end{aligned}$$

where $|H| < 1/4$. The right member of (8.8) is of the form $b + |b|c$. Write $b + |b|c - |b|c = b$. Then $|(b + |b|c)| + |b||c| \geqq |b|$, or $|(b + |b|c)| \geqq |b|(1 - |c|)$. Hence

$$(8.9) \qquad |a(n_i t_i)| \geqq |n_i a(t_i)|(1 - |H|) > \tfrac{3}{4}|n_i a(t_i)| = \tfrac{3}{4} n_i t_i \left|\frac{a(t_i)}{t_i}\right|.$$

Since $n_i t_i \to 1$ and $a(t_i)/t_i > 2|a|$, the last expression in (8.9) becomes and remains greater than $5/4|a|$ when i is large. This is impossible since $a(n_i t_i) \to a$.

(8.10) If $O \neq a \in U$, $\lim a(t)/t$ ($t \to 0$) exists and is different from O.

Proof. By virtue of the preceding proposition it is sufficient to prove that if $\{s_i\}$, $\{t_i\}$ are sequences of positive numbers converging to 0 and such that $\lim a(s_i)/s_i \to A$, $\lim a(t_i)/t_i \to B$, then $A = B$. On suppressing a number of the first terms in the sequence $\{s_i\}$ if necessary, we may suppose s_1 to be so small that $|a(s_1)/s_1 - A| < |A - B|/3$ and that

$$(8.11) \qquad |F(a(t), a(t'))| < |A - B|/3$$

when $0 \leqq (t, t') \leqq s_1$. For each t_i choose an integer n_i such that

$$s_1\left(1 - \frac{n_i}{n_i + 1}\right) \leqq n_i t_i \leqq s_1.$$

Then $n_i t_i \to s_1$ so that $|a(n_i t_i) - a(s_1)| \to 0$. Hence we may write

$$(8.12) \qquad |a(n_i t_i) - s_1 A| < \frac{s_1}{3}|A - B| + \epsilon_i$$

where $\epsilon_i \to 0$. Use of (8.8) gives

$$a(n_i t_i) = n_i t_i \left(\frac{a(t_i)}{t_i} + \left| \frac{a(t_i)}{t_i} \right| H \right).$$

On examining the definition of H in (8.8), we see that (8.11) implies here that $| H | < | A - B | /3$. Since $n_i t_i \to s_1$ and $a(t_i)/t_i \to B$, it follows that

$$| a(n_i t_i) - s_1 B | = \frac{s_1}{3} | A - B | + \eta_i$$

where $\eta_i \to 0$ as $i \to \infty$. Combine this with (8.12) to obtain

$$s_1 | A - B | \leqq \frac{2s_1}{3} | A - B | + \epsilon_i + \eta_i .$$

We conclude that $| A - B | \leqq (2/3) | A - B |$ which implies that $A = B$.

9. Canonical mappings

We continue to assume that the coordinate system for G_r is left-differentiable. From now on we shall denote $a(t)$ by a^t. For each a we extend the function a^t over the interval $(-1, +1)$ by the formula $a^{-t} = (a^t)^{-1}$. The function a^t $(-1 \leqq t \leqq 1)$ is defined for every a in a certain spherical nucleus $U \subset W$. By (8.6) we may assume U to be so small that $a^t a^s$ and $(a^t)^s$ are defined and in W when $a \in U$ and $| t |, | s | \leqq 1$.

(9.1) Let a be an element in U. Then $a^s a^t = a^{s+t}$ when $| s |, | t |, | s + t | \leqq 1$ and $(a^s)^t = a^{st}$ when $| s |, | t | \leqq 1$. In canonical coordinates, $a^t = ta$.

PROOF. The first formula holds of course when s, t are positive. Extension to the interval $(-1, +1)$ is an immediate consequence of the relation $a^{-t} = (a^t)^{-1}$. Similarly it will be sufficient to prove the second formula for positive t, s. Since $a^{t/2} \in W \subset Z$, we have $| a^{t/2} | \leqq | (a^{t/2})^2 | = | a^t |$. Hence $a^{t/2}$ is a principal square root of a^t. Since $a^t \in W$, $a^{t/2}$ is the *unique* principal square root of a^t, and hence equal to $(a^t)^s$ with $s = 1/2$. (See proof of (8.6)). Thus the formula in question holds when $s = 1/2$. Similarly it holds when $s = 1/2^n$, then for $s = m/2^n$, hence, by continuity, for arbitrary s.—The relation $a^t = ta$ for canonical coordinates follows immediately from the proof of (8.6).

(9.2) a^t converges to O uniformly with respect to $t(-1 \leqq t \leqq 1)$ when $a \to O$. For, by (8.6), $a \in V_\epsilon$ implies $a^t \in V_{8\epsilon} + V_{8\epsilon}^{-1}$.

(9.3) The t-derivative of a^t exists at $t = 0$. For, (8.10) implies that a^t has a right derivative at $t = 0$. It is a consequence of (3.13) that for $t \to +0$, $\lim a^{-t}/t = \lim (a^t)^{-1}/t = -\lim a^t/t$. Hence the left derivative exists at $t = 0$ and equals the right derivative.

We now note certain properties of the function

$$\Lambda(a) = \frac{da^t}{dt_0} = \lim_{t \to 0} \frac{a^t}{t} = \lim_{n \to \infty} n a^{\frac{1}{n}} .$$

(9.4) $\Lambda(a)$ is defined over U and $| \Lambda(a) | \leqq 2 | a |$. (This inequality is a consequence of (8.7).)

Left differentiability of ab and differentiability of a^s at $s = 0$ imply the differentiability of $a^s b$ with respect to s at $s = 0$ according to the formula

$$\frac{d}{ds_0} a^s b = \sum \left(\frac{\partial(ab)}{\partial a^i}\right)_{a=0} (\Lambda(a))^i = \sum (\Lambda(a))^i \psi_i(b)(3.1).$$

Putting $b = a^t$ we obtain

$$(9.5) \qquad \sum (\Lambda(a))^i \psi_i(a^t) = \lim_{s \to 0} \frac{1}{s} (a^s a t - a^t) = \lim_{s \to 0} \frac{1}{s} (a^{s+t} - a^t) = \frac{da^t}{dt}$$

so that a^t *is t-differentiable throughout.*

(9.6) There exists a function $M(a)$ which is continuous at $a = O$ with $M(O) = O$ and is such that

$$\Lambda(a) = a + |\Lambda(a)| M(a).$$

Proof. We have only to show that $(\Lambda(a) - a)/|\Lambda(a)| \to O$ as $a \to O$. From (9.5) we have

$$(t\Lambda(a) - a^t)^i = \sum_j (\Lambda(a))^j \int_0^t \delta_j^i - \psi_j^i(a^\tau)\, d\tau.$$

Hence

$$(9.7) \qquad \frac{|(t\Lambda(a) - a^t)_j^i|}{|\Lambda(a)|} \leqq \sum_j \int_0^t |\delta_j^i - \psi_j^i(a^\tau)|\, d\tau \leqq t \sum_j M_j^i$$

where

$$M_j^i = \max \{|\delta_j^i - \psi_j^i(a^\tau)|, 0 \leqq \tau \leqq t\}.$$

Since $\psi(b)$ is locally continuous and $\psi_j^i(O) = \delta_j^i$ it follows from (9.2) that for given t, $M_j^i \to 0$ as $t \to 0$. The limit in question is now verified by putting $t = 1$ in (9.7) and letting $a \to 0$.

Certain useful formulas come from (9.5). We have

$$|\Lambda(a)| \leqq |a| + |\Lambda(a)||M(a)|$$

so that $|\Lambda(a)|/|a| \leqq (1 - |M(a)|)^{-1}$. Hence

$$\frac{|\Lambda(a)|}{|a|} = 1 + \rho(a)$$

where $\rho(a)$ has real values and converges to 0 as $a \to O$. From (9.6) we now obtain

$$\frac{\Lambda(a)}{|a|} = \frac{a}{|a|} + |(1 + \rho(a))| M(a)$$

which, on multiplication by $|a|$ becomes

$$(9.8) \qquad \Lambda(a) = a + |a| M'(a)$$

where $M'(a) \to O$ as $a \to O$.

The formula

$$\Lambda(a^t) = t\Lambda(a) \tag{9.9}$$

is a consequence of (9.1). For,

$$\Lambda(a^t) = \lim_{s \to 0} \frac{(a^t)^s}{s} = t \lim_{s \to 0} \frac{a^{ts}}{ts} = t\Lambda(a).$$

(9.10) $\Lambda(a)$ is locally continuous.

PROOF. The correspondence $a \to a^{1/2}$ is single-valued over $\bar{U}$ (since $\bar{U} \subset W$, see (8.5)) and has a single-valued inverse, namely $a \to a^2$. Since this last correspondence is continuous and $\bar{U}$ is compact, the correspondence $a \to a^{1/2}$ is itself continuous. Hence the functions $a^{1/2}$, $a^{1/4}$, $\cdots$ are continuous over U. Now from (9.9) and (9.6),

$$\begin{aligned} \Lambda(a) = n\Lambda(a^{1/n}) &= na^{1/n} + |\Lambda(a)| M(a^{1/n}) \\ &= na^{1/n} + H(a, n) \quad \text{say.} \end{aligned} \tag{9.11}$$

We assert that $H(a, n)$ converges to O uniformly with respect to a as n converges to ∞ through powers of 2. This is a consequence of the relations: $|\Lambda(a)| \leqq 2|a|$ (see 9.4) and $|a^{1/2^n}| \leqq (3/4)^n |a|$ (8.2).—From (9.11) we have

$$|\Lambda(a) - \Lambda(b)| \leqq n|a^{1/n} - b^{1/n}| + |H(a, n) - H(b, n)|. \tag{9.12}$$

Let ϵ be a positive number. Take for n such a large power of 2 that the last term in (9.12) is smaller than $\epsilon/2$. Since $a^{1/n}$ is continuous, the first term in the right member of (9.12) becomes smaller than $\epsilon/2$, hence $|\Lambda(a) - \Lambda(b)|$ becomes smaller than ϵ when $|a - b|$ is sufficiently small.

Another formula:

$$\lim_{t \to 0} \frac{1}{t} \Lambda(a^t b^t) = \Lambda(a) + \Lambda(b). \tag{9.13}$$

PROOF. Writing $c_t = a^t b^t$, we note that $\lim c_t/t = \Lambda(a) + \Lambda(b)$. This follows from the relation $c_t = a^t + b^t + |a^t| F(a^t, b)$ (3.7). The formula in question now comes from using (9.8) with c_t replacing a.

We shall say that the mapping $\kappa: a \to \Lambda(a)$ is *canonical* if its inverse is single-valued.

Suppose κ is canonical. Then if U' is any nucleus whose closure is contained in U, κ maps U' homeomorphically and $\kappa(U')$ is an open set containing O. Write $\mathbf{a}^i = \Lambda^i(a)$ and regard the $\mathbf{a}^i$ as coordinates in a space $\mathbf{G}_r$. If we introduce products and inverses into $\mathbf{G}_r$ by the formulas $\Lambda(a)\Lambda(b) = \Lambda(ab)$, $(\Lambda(a))^{-1} = \Lambda(a^{-1})$, then $\mathbf{G}_r$ becomes a local group, κ a local isomorphism $G_r \to \mathbf{G}_r$. Moreover, (9.9) and (9.13) imply that the coordinates $\mathbf{a}^i$ are canonical. Hence

(9.14) *If κ is canonical, G_r is locally isomorphic to a local group with canonical coordinates.*

10. Vector fields

Let R be an open set in a euclidean space S_n with coordinates $x^1, \cdots, x^n$. A vector field ξ over R is a single-valued function $\xi(x)$ or ξx defined over R, with values in S_n. To say that ξ is of class k over R means that the k-fold derivatives of the functions $\xi^i(x^1, \cdots, x^n)$ exist and are continuous over R. We shall consider also the extreme cases in which ξ is analytic, or merely continuous, over R.

Let ξ be a continuous vector field over R. For each point x in R there exists at least one solution $f(x, t)$ of the system of equations $dx^i/dt = \xi^i(x)$ satisfying the initial condition $f(x, 0) = x$, and defined for all t in a neighborhood of $t = 0$. Suppose that for every choice of x in R the solution $f(x, t)$ is unique in the sense that if $g(x, t)$ is also a solution with $g(x, 0) = x$, then $g(x, t) = f(x, t)$ for all values of t in some neighborhood of $t = 0$ depending on x. Then for given x, the solution curve through x can be continued in both directions in a unique manner, either indefinitely or until $f(x, t)$ runs out of R. Thus for given x, $f(x, t)$ is defined and single-valued over an interval $-t_1 < t < t_2$ where t_1, t_2 are positive and one or both may be ∞. We shall say that ξ possesses the *uniqueness property* over R and shall denote $f(x, t)$ by $e^{t\xi}x$. The solution $e^{t\xi}x$ is additive in t: $e^{t\xi}e^{s\xi}x = e^{(t+s)\xi}x$. By standard existence theorems, vector fields which are analytic or of class k, $k \geqq 1$, possess the uniqueness property.

Let $\xi_1, \cdots \xi_r$ be vector fields of class $k \geqq 1$ over R and let $a_1, \cdots, a^r$ be real variables. Let R' be a bounded open set such that $\bar{R}' \subset R$ and let K be a positive number. There exists a number $\delta > 0$ such that $e^{t(a^1\xi_1 + \cdots + a^r\xi_r)}x$ is defined for $x \in R'$, $|t| < \delta$, $|a| < K$ $(a = a^1, \cdots, a^r)$. Hence $e^{a^1\xi_1 + \cdots + a^r\xi_r}x (t = 1)$ is defined for $x \in R'$ and $|a| < \delta K$. Moreover, as a function of $a^1, \cdots, a^r$, $x^1, \cdots, x^n$, $e^{a^1\xi_1 + \cdots + a^r\xi_r}x$ is of class k; it is analytic if the ξ_i are analytic. For each a with $|a| < \delta K$, the transformation $x \to e^{a^1\xi_1 + \cdots + a^r\xi_r}x$ transforms R' homeomorphically.

Suppose that ξ is of class 1. Then

$$\text{(10.1)} \qquad \frac{d}{dt} e^{t\xi}x = \xi(e^{t\xi}x); \qquad \frac{d}{dt_0} e^{t\xi}x = \xi(x);$$

$$\text{(10.2)} \qquad \frac{d}{dt}\frac{\partial}{\partial x^i} e^{t\xi}x = \frac{\partial}{\partial x^i}\xi(e^{t\xi}x) = \sum_j \left(\frac{\partial \xi(y)}{\partial y^j}\right)_{y=e^{t\xi}x} \frac{\partial (e^{t\xi}x)^j}{\partial x^i}.$$

Formulas (10.1) are implied by the definitions they involve. Formula (10.2) is what we obtain by differentiating first with respect to t, then with respect to x^i. Since the resulting function is continuous in (t, x) and since the same is true of the x^i-derivative of $e^{t\xi}x$ it follows from a standard theorem that the order of differentiation can be reversed.

We shall say that the vector fields $\xi_1, \cdots, \xi_r$ are *linearly independent* over R if there exist no real numbers $c_1, \cdots, c_r$ not all zero and no open subset R' of R such that $\sum c_i\xi_i(x) = O$ identically over R'.

11. Realizations

The continuous groups of the classical theory are transformation groups. From our point of view they may be regarded as realizations of local groups by means of transformations.

DEFINITION. *Let $G = \{\mathbf{1}, a, b, \cdots\}$ be a local group, $S = \{x, y, \cdots\}$ a space. Suppose there exists a function $a \cdot x$ ($a \in G$, $x \in S$) which is single-valued and continuous in the pair (a, x) wherever defined, has points of S as values, and satisfies the following conditions:* (1) *whenever defined, $\mathbf{1} \cdot x$ equals x;* (2) *if $a \cdot (b \cdot x)$, $(ab) \cdot x$ are defined, they are equal;* (3) *$a \cdot x$ equals $a \cdot y$ only if $x = y$;* (4) *there is a nucleus U of G and an open set R in S such that $U \cdot R$ is defined (i.e. $a \cdot x$ is defined when $a \in U$, $x \in U$). We then say that there is defined a realization (G, S) of G. A realization (G, S) is effective if there exists a nucleus U_0 and open set R_0 in S such that $U_0 \cdot R_0$ is defined and such that the only element a in U_0 such that $a \cdot x \equiv x$ over R_0 is $a = \mathbf{1}$. For every local group G there is an effective realization (G, S) namely the regular realization obtained by taking $G = S$ and defining $a \cdot b$ to be ab.*

The following theorem and proof (Lie) are classical. We add here an element of precision required for our purposes.

(11.1) Let G_r, S_n be euclidean spaces of dimensions r, n with coordinate systems $a^1, \cdots, a^r$ and $x^1, \cdots, x^n$. Let U be a neighborhood of the origin of G_r, R an open subset of S_n. Let $\phi_1, \cdots, \phi_r$ be vector fields of class k ($k \geqq 1$) over U such that the matrix $(\phi_j^i(O))$ is non-singular. Let $\xi_1, \cdots, \xi_r$ be linearly dependent set of continuous vector fields over R such that every vector field $\sum a^i \xi_i$ possesses the uniqueness property over R. Let $a \cdot x$ be a single-valued function defined for $a \in U$, $x \in R$ with values in S_n and such that $O \cdot x = x$ when $x \in R$. Assume that the derivatives $\partial(a \cdot x)^i / \partial a^j$ exist and that over U, R,

$$\frac{\partial (a \cdot x)^i}{\partial a^j} = \sum_h \xi_h^i(a \cdot x) \phi_j^h(a). \tag{11.1a}$$

Then there exists a function ab which converts G_r into a local r-parameter group for which $a \cdot x$ defines an effective realization (G_r, S_n). The group-theoretic structure of G_r is determined solely by the vector fields ϕ_i. The coordinate system $a^1, \cdots, a^r$ of G_r is of class k. Moreover G_r is locally isomorphic to a local group $\mathbf{G}_r$ with canonical coordinates $\mathbf{a}^1, \cdots, \mathbf{a}^r$ of class k. The functions

$$\mathbf{a} * x = e^{\mathbf{a}^1 \xi_1 + \cdots \mathbf{a}^r \xi_1} x \tag{11.2}$$

define an effective realization $(\mathbf{G}_r, S_n)$. The coordinate systems of G_r and $\mathbf{G}_r$ are analytic if the ϕ_i are analytic.

PROOF. Let ψ be the inverse of the matrix $\phi = (\phi_j^i(a))$. Evidently $\psi(a)$ exists and is of class k in the neighborhood of $a = O$. Let $\mathbf{a}^1, \cdots, \mathbf{a}^r$ be coordinates in a space $\mathbf{G}_r$ and let

$$\chi(\mathbf{a}) = e^{\Sigma \mathbf{a}^i \psi_i} O.$$

We assert that the formula $a = \chi(\mathbf{a})$ defines a homeomorphism of class k be-

tween neighborhoods of the origins of G_r and $\mathbf{G}_r$. To prove this, it is sufficient to show that the matrix $(\partial a^i/\partial \mathrm{a}^j)$ is non-singular at $\alpha = O$. We have

$$\left(\frac{\partial a^i}{\partial \mathrm{a}^j}\right)_O = \left(\frac{d}{d\mathrm{a}^j}\, e^{\mathrm{a}^j\psi_j}O\right)^i_{\mathrm{a}^j=0} = \left(\frac{d}{dt_0}\, e^{t\psi_j}O\right)^i = \psi_j^i(O).$$

We introduce products ab and $\mathbf{ab}$ into G_r and $\mathbf{G}_r$ by the formulas

$$ab = e^{\Sigma \mathbf{a}^i\psi_i}e^{\Sigma \mathbf{b}^i\psi_i}O \quad \text{where} \quad \mathbf{a} = \chi^{-1}(a), \quad \mathbf{b} = \chi(b);$$

$$\mathbf{ab} = \chi^{-1}(ab).$$

We shall see eventually that these functions convert G_r, $\mathbf{G}_r$ into locally isomorphic local groups, and that (11.2) defines an effective realization of $\mathbf{G}_r$. The structure of G_r is determined by the ψ_i hence by the ϕ_i. The functions ab and $\mathbf{ab}$ are evidently of class k or analytic according as the ϕ_i are of class k or analytic. Moreover the coordinates $\mathbf{a}^i$ will be canonical since

$$(s\mathbf{a})(t\mathbf{a}) = \chi^{-1}(e^{s\Sigma \mathbf{a}^i\psi_i}e^{t\Sigma \mathbf{a}^i\psi_i}O) = \chi^{-1}(e^{(s+t)\Sigma \mathbf{a}^i\psi_i}O) = (s+t)\mathbf{a}.$$

This shows incidentally that the origin of $\mathbf{G}_r$ plays the part of the identity and that elements near it have inverses: $\mathbf{a}^{-1} = -\mathbf{a}$. Therefore to show that $\mathbf{G}_r$ is a local group it is sufficient to show that the associative law holds in the neighborhood of O. $\mathbf{G}_r$ will then also be a local group locally isomorphic to G_r thru the relation $ab = \chi(\mathbf{ab})$.

We first establish the identity

$$\mathbf{a}*(b\cdot x) = (e^{\Sigma \mathbf{a}^i\psi_i}b)\cdot x \tag{11.3}$$

which holds for x in R and $|\,\mathbf{a}\,|$, $|\,b\,|$ sufficiently small depending on x. If we replace $\mathbf{a}^i$ by $t\mathbf{a}^i$ in (11.3), the left member is that solution of the system $dx/dt = \sum \mathbf{a}^i\xi_i$ which equals $b\cdot x$ when $t = 0$. But so is the right member. For if we write $b_t = e^{t\Sigma \mathbf{a}^i\psi_i}b$ and use the fact that $\sum_j \phi_j^h(b)\psi_i^j(b) = \delta_i^h$ we obtain

$$\frac{d}{dt}(b_t\cdot x)^k = \sum_j \left(\frac{\partial(a\cdot x)^i}{\partial a^j}\right)_{a=b_t}\frac{db_t^j}{dt} = \sum_{ijh}\mathbf{a}^i\xi_h^k(b_t\cdot x)\phi_j^h(b_t)\psi_i^j(b_t) = \sum_h \mathbf{a}^h\xi_h^k(b_t\cdot x).$$

Since $\sum \mathbf{a}^j\xi_j$ possesses the uniqueness property we conclude that (11.3) holds with the $\mathbf{a}^i$ are replaced by $t\mathbf{a}^i$. In particular we obtain (11.3) by putting $t = 1$.

Let $a = \chi(\mathbf{a})$, $b = \chi(\mathbf{b})$. On putting $b = O$, (11.3) becomes

$$\mathbf{a}*x = \chi(\mathbf{a})\cdot x = a\cdot x. \tag{11.4}$$

Then on replacing x by $b\cdot x$ in (11.4) we obtain with (11.3):

$$a\cdot(b\cdot x) = \mathbf{a}*(b\cdot x) = (e^{\Sigma \mathbf{a}^i\psi_i}b)\cdot x = (e^{\Sigma \mathbf{a}^i\psi_i}e^{\Sigma \mathbf{b}^i\psi_i}O) = ab\cdot x. \tag{11.5}$$

By (11.4) we have equally well $\mathbf{a}*(\mathbf{b}*x) = \mathbf{ab}*x$. This relation and (11.5) hold for x in R and for $|\,a\,|$, $|\,b\,|$, $|\,\mathbf{a}\,|$, $|\,\mathbf{b}\,|$ sufficiently small depending on x. It

will be seen, in fact, that if R' is a bounded open set such that $\bar{R}' \subset R$ the relations in question, namely

$$(11.6) \qquad a\cdot(b\cdot x) = ab\cdot x, \qquad \mathbf{a}*(\mathbf{b}*x) = \mathbf{ab}*x,$$

hold for $x \in R'$ and $|a|, |b|, |\mathbf{a}|, |\mathbf{b}| < \beta$ say, where β is independent of x.

We assert next that there exists a positive number γ such that if $|\mathbf{a}| < \gamma$ and if $\mathbf{a}*x = x$ identically over R', then $\mathbf{a} = 0$. If this were not the case, there would exist a sequence $\mathbf{a}_i \to 0$ such that $\mathbf{a}_i*x \equiv x$ over R'. Then since $x = \mathbf{a}_i^n*x = n(\mathbf{a}_i*x)$, we conclude by an obvious limiting process (cf. proof of (5.7) that $(t\mathbf{e})*x \equiv x$ over R' for every t with $|t|$ sufficiently small, $\mathbf{e}$ being a suitable unit vector. From (11.2) we have $e^{t(\mathbf{e}^1\xi_1+\cdots+\mathbf{e}^r\xi_r)}x \equiv x$ over R' which can only happen if $\sum \mathbf{e}^i\xi_i(x) \equiv 0$ over R' which is impossible since the ξ_i are linearly independent over R.

It follows from this that if $\mathbf{a}*x \equiv \mathbf{b}*x$ over $R'(\mathbf{a}, \mathbf{b}$, near $O)$, then $\mathbf{a} = \mathbf{b}$. For, $(\mathbf{b}^{-1}\mathbf{a})*x \equiv \mathbf{b}^{-1}(\mathbf{a}*x) \equiv x$, hence $\mathbf{b}^{-1}\mathbf{a} = O$, $\mathbf{b} = \mathbf{a}$. From this comes the associative law. For if $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are near O, then $(\mathbf{a}(\mathbf{bc}))*x \equiv ((\mathbf{ab})\mathbf{c})*x$ over R' since both sides can be reduced to $\mathbf{a}*(\mathbf{b}*(\mathbf{c}*x))$ by applications of (11.6). Hence $(\mathbf{ab})\mathbf{c} = \mathbf{a}(\mathbf{bc})$.—Thus $\mathbf{G}_r$ is a local group for which $\mathbf{a}*x$ defines a realization. That this realization is effective follows from the preceding paragraph. As we have already remarked, the function ab converts G_r into a local group locally isomorphic with $\mathbf{G}_r$. From the relations $a\cdot x = \mathbf{a}*x$ (11.4) it follows that the functions $a\cdot x$ define an effective realization of G_r.

12. The Lipschitz condition

We shall say that a coordinate system for G_r *satisfies a right Lipschitz condition* if there exists a nucleus V and a positive number C such that $|ab - ac| < C|b - c|$ for $a, b, c \in V$.

(12.1) *In a left differentiable coordinate system which satisfies a right Lipschitz condition, the mapping $a \to \Lambda(a)$ (§9) is canonical. Moreover, there exists a nucleus over which the vector fields ψ_i (3.1) possess the uniqueness property.*

PROOF. Let a be an element in the nucleus U (domain of definition of $\Lambda(a)$) and consider the system

$$(12.2) \qquad \frac{df}{dt} = \sum \mathbf{a}^i\psi_i(f), \qquad (\mathbf{a}^1, \cdots, \mathbf{a}^r) = \mathbf{a} = \Lambda(a).$$

Formula (9.5) shows that a^t is a solution of (12.2). We assert moreover that if b is an element in U, then a^tb is also a solution. For

$$(12.3) \qquad \frac{d}{dt}a^tb = \lim_{s\to 0}\frac{1}{s}(a^{t+s}b - a^tb)$$

$$= \lim_{s\to 0}\frac{1}{s}(a^s(a^tb) - a^tb) = \frac{d}{ds_0}a^s(a^tb) = \sum \mathbf{a}^i\psi_i(a^tb).$$

Now suppose that $f(t)$ is a solution which assumes the same initial value as a^tb—namely b—when $t = 0$. We shall show that $f(t) = a^tb$, or what amounts

to the same thing, that the function $g(t) = (a^t b)^{-1} f(t)$ equals O for all values of t sufficiently near $t = 0$. For this it is sufficient to show that the derivative of $g(t)$ exists and equals O identically in the néighborhood of $t = 0$. We have

$$\Delta g = | g(t + s) - g(t) | = | b^{-1}a^{-t-s}f(t + s) - b^{-1}a^{-t-s}a^s f(t) |$$
$$< C \,| f(t + s) - a^s f(t) |$$

provided $| t |$ and $| s |$ are not too large. From (12.2) we have

$$\frac{d}{ds_0}(f(t + s) - a^s f(t)) = \sum \mathbf{a}^i \psi_i(f(t)) = \sum \left(\frac{\partial af(t)}{\partial a^i}\right)_{a=0} \frac{d(a^s)^i}{ds_0} = O.$$

Hence $| \Delta g | < C \,| h | \,\epsilon\,(s, t)$ say where $\epsilon(s, t) \to 0$ as $s \to 0$. Hence $\Delta g/s \to O$ and therefore $dg/dt = O$.

Suppose in particular that $b = O$. Then any solution of (12.2) which equals O when $t = 0$ is identical with a^t near $t = 0$. Suppose now that $\Lambda(c) = \Lambda(a)$. Then c^t also satisfies (12.2) and hence $c^t = a^t$ in a neighborhood of $t = 0$. We conclude (with the aid of (9.1)) that $c^t = a^t$ throughout, hence that $c = a$. It follows that the mapping $a \to \Lambda(a)$ is canonical.

We have shown that $\sum \mathbf{a}^i\psi_i$ possesses the uniqueness property over U when $\mathbf{a}$ is the image of an element a under the mapping $a \to \Lambda(a)$. Since this mapping is canonical, we have established uniqueness when a is any element in a certain neighborhood of O. It is easy to see however that if uniqueness holds for $\sum \mathbf{a}^i\psi_i$, then it holds also for $\lambda(\sum \mathbf{a}^i\psi_i)$ where λ is a real number. Hence for arbitrary a, $\sum \mathbf{a}^i\psi_i$ possesses the uniqueness property over U.

We have still to show that the ψ_i are linearly independent. From (12.3) with $t = 1$, we have

$$ab = e^{\Sigma \mathbf{a}^i \psi_i} b$$

over a sufficiently small nucleus V. If the ψ_i are not linearly independent over V, we may suppose that $\psi_r = c_1\psi_1 + \cdots + C_{r-1}\,\psi_{r-1}$ identically over some open subset V' of V. Then for every a such that $a = (0, \cdots, 0, a^r)$ ($| a^r |$ small) we have $ab = b$ which is impossible.

THEOREM (12.4) (van Kampen).* *A G_r with a left-differentiable coordinate system satisfying a right Lipschitz condition is locally isomorphic to a local r-parameter group with analytic canonical coordinates.*

Let $x^1, \cdots, x^r$ be the coordinate system of G_r. By (12.1) the mapping $x \to \mathbf{x} = \Lambda(x)$ is canonical and hence it establishes a local isomorphism between G_r and a local group $\mathbf{G}_r$ with canonical coordinates $\mathbf{x}^i$. Let x, y be elements in the nucleus U (existence domain of $\Lambda(x)$) and let $\mathbf{x} = \Lambda(x)$, $\mathbf{y} = \Lambda(y)$. Since $\mathbf{x}^t = t\mathbf{x}$, we have (7.1 and 7.3):

$$(t\mathbf{x})\,(s\mathbf{y})\,(-t\mathbf{x}) = s\,(e^{-t\mathbf{X}}\mathbf{y}) = s\mathbf{y}', \quad \text{say}$$

where X is the linear transformation $\mathbf{z} \to [\mathbf{x}, \mathbf{z}]$. Hence (9.9), $x^t y^s x^{-t} = y'^s$ where $y' = \Lambda^{-1}(\mathbf{y})$. Thus

$$(12.5) \qquad x^t y^s = y'^s x^t \quad \text{where} \quad \mathbf{y}' = e^{-tX}\mathbf{y}, \qquad y = \Lambda(\mathbf{y}).$$

Now let $y = y_1^{a^1} \cdots y_r^{a^r}$ where the y_i are fixed elements in U and the a^i are real variables. Let x be an element in U. Then

$$\frac{\partial(yx)}{\partial a^p} = \lim_{s \to 0} \frac{1}{s} (y_1^{a^1} \cdots y_p^{a^p+s} \cdots y_r^{a^r} x - yx).$$

Write $y^s y_p^{a^p}$ in place of y^{a^p+s}, then transfer y^s to the left, step by step using (12.5). We obtain

$$\frac{\partial(yx)}{\partial a^p} = \lim_{s \to 0} \frac{1}{s} z_p^s yx - yx$$

where

$$\Lambda z_p = \mathbf{z}_p = e^{-a^1 X_1} \cdots e^{-a^{p-1} X_{p-1}} \mathbf{y}$$

in which X_i is the linear transformation $\mathbf{w} \to [\mathbf{y}_i, \mathbf{w}]$, $\mathbf{y}_i = \Lambda(y_i)$. Hence

$$(12.6) \qquad \frac{\partial(yx)}{\partial a^p} = \frac{d}{ds_0} z_p^s yx = \sum_i \frac{\partial w(yx)}{\partial w^i} \frac{d(z_p^s)^i}{ds_0} \sum (\mathbf{z}_p)^i \psi_i(yx).$$

Now if λ is a small enough positive number, the elements y_j can be so chosen that $y_j^i = \lambda\delta_j^i$; let the y_j be so fixed. Then y is a function θ of $a^1, \cdots, a^r$ alone. Moreover the matrices of the transformations X_i are constants and hence $\mathbf{z}_p$ is a convergent power series ϕ_p in $a^1 \cdots, a^{p-1}$. If, in the equations

$$(12.7) \qquad \frac{\partial(yx)}{\partial a^p} = \sum_i \psi^i(yx)\phi_i^p(a)$$

which (12.6) now become, we put $x = O$, we see that the derivatives $\partial y/\partial a^j$ exist and are continuous. The matrix $(\partial y^i/\partial a^j)$ is non-singular at $a = O$. In fact

$$\left(\frac{\partial y^i}{\partial a^j}\right)_{a=O} = \frac{d}{dt_0}(y_j^t)^i = \Lambda(y_j)^i = \lambda\delta_j^i.$$

It follows that the correspondence $a \to \theta(a) = y$ maps a nucleus of G_r homeomorphically onto a neighborhood of the origin of a space G_r' with coordinates $a^1, \cdots, a^r$. Let us define a function $a\cdot x$ by the formula $a\cdot x = yx$, $y = \theta(a)$. Then (12.7) becomes

$$\frac{\partial(a\cdot x)}{\partial a^h} = \sum \psi^i(a\cdot x)\phi_i^h(a).$$

Since the ψ_i are linearly independent and possess the uniqueness property over some nucleus of G_r (12.1), there exists (11.1) a product function ab converting

G'_r into a local group for which $a^1, \cdots, a^r$ is an analytic coordinate system. Since (11.1) G'_r is locally isomorphic to a local group with analytic canonical coordinates, it remains only to be shown that G'_r is locally isomorphic to G_r. The correspondence $a \rightarrow \theta(a)$ is in fact such an isomorphism. For,

$$(\theta(ab))x = ab \cdot x = (\theta(a))\,(b \cdot x) = (\theta(a)\theta(b))x$$

identically in x. Hence $\theta(ab) = \theta(a)\,\theta(b)$ which completes the proof.

Let us note that according to (11.1) the structure of G'_r (hence that of G_r) is determined by the vector fields ϕ_i. These in turn are determined solely by the transformations X_i hence, ultimately by the function $[\mathbf{x}, \mathbf{y}]$ over $\mathbf{G}_r$. Hence if we take $G_r = \mathbf{G} =$ a local group with analytic canonical coordinates, we obtain

THEOREM (12.8). *The group-theoretic structure of a r-parameter local group with an analytic canonical coordinate system is completely determined by its commutator function.*

THEOREM (12.9). *Every left-differentiable canonical coordinate system satisfying a right Lipschitz condition is analytic. Hence in such a coordinate system the vector fields ψ_i are analytic. Products are given by the formula*

$$ab = e^{\Sigma a^i \psi_i} b \qquad (a = a^1, \cdots, a^r). \tag{12.10}$$

PROOF. Let G_r have a coordinate system Σ with the stated properties. G_r is locally isomorphic to G'_r with analytic canonical coordinates Σ'. By (5.4) the isomorphism $G_r \rightarrow G'_r$ is linear in terms of Σ and Σ'. Hence Σ is analytic.—Since Σ is canonical so that $a^t = ta$ (9.1), the canonical mapping applied to G_r is the identity; that is, $a = \Lambda(a) = a$. Hence (12.3) becomes

$$\frac{d}{dt}\, a^t b = \Sigma a^i \psi_i(a^t b)$$

which implies (12.10).

13. Vector fields of class 1

Let ξ, η be vector fields of class 1 over an open set R in a space S_n with coordinate system $x^1, \cdots, x^n$. The continuous vector field

$$[\xi,\eta]^i x = \sum_j \left(\xi^j(x) \frac{\partial \eta^i(x)}{\partial x^j} - \eta^j(x) \frac{\partial \xi^i(x)}{\partial x^j} \right)$$

is the *commutator* of ξ, η. Evidently

$$[\xi, \eta] = -[\eta, \xi]. \tag{13.1}$$

If ξ, η, ζ, $[\xi, \eta]$, $[\eta, \zeta]$, $[\zeta, \xi]$ are of class 1, then

$$[\xi, [\eta, \zeta]] + [\eta, [\zeta, \xi]] + [\zeta, [\xi, \eta]] = O \tag{13.2}$$

The verification of this well-known identity is straightforward.

We now study the transformation

$$T_{t,s} = e^{-t\xi} e^{s\eta} e^{t\xi} x \tag{13.3}$$

where ξ, η are of class 1 over R. Let $z = e^{s\eta} e^{t\xi}x$, $y = e^{t\xi}x$. Then since $z = y$ when $s = 0$,

$$(13.4) \qquad \left(\frac{d}{ds_0} T_{t,s} x\right)^i = \left(\frac{d}{ds_0} e^{-t\xi} z\right)^i = \sum_j \left(\frac{\partial(e^{-t\xi} w)^i}{\partial w^j}\right)_{w=y} \eta^j(y).$$

We could equally well have computed the s-derivative for arbitrary s. It is sufficient for our purposes to note that this derivative exists.

We assert that the t-derivative of dT/ds_0 exists. The second factor in the right of (13.4) possesses a t-derivative and it is sufficient to show therefore that the first factor does also. The existence of this derivative is a consequence of the relation

$$(13.5) \qquad \delta_k^i = \frac{\partial}{\partial x^k} (e^{-t\xi} e^{t\xi} x)^i = \sum_j \left(\frac{\partial(e^{-t\xi} w)}{\partial w^j}\right)^i_{w=y} \frac{\partial y^j}{\partial x^k}$$

and the fact that the t-derivative of $\partial y^j/\partial x^k$ exists (see (10.2)).

We obtain an explicit formula for the t-derivative of dT/ds_0 as follows. Let

$$(13.6) \qquad T^i = \frac{d(Tx)^i}{ds_0}, \qquad B_j^i = \left(\frac{\partial(e^{-t\xi} w)^i}{\partial w^j}\right)_{w=y}, \qquad C^j = \eta^j(y), \qquad D_k^j = \frac{\partial y^j}{\partial x^k}.$$

Formula (13.4) now becomes

$$(13.7) \qquad \frac{dT^i_{s,t}}{ds_0} = T^i = \sum B_j^i C^j = \sum B_j^i \eta^j(y)$$

and (13.5) becomes $\delta_k^i = \sum B_j^i D_k^j$. Since the matrices B, D are inverses of each other, so are their transposes. Hence $\delta_j^k = \sum B_j^i D_i^k$. Combining this with (13.7) we obtain $C^k = \sum T^i D_i^k$. Now take the t-derivative of this last relation, multiply both sides by B_k^j, sum with respect to k and use the relation $\sum B_k^j D_i^k = \delta_i^j$. The result after a transposition is

$$(13.7\text{a}) \qquad \frac{dT^j}{dt} = \sum_k B_k^j \left(\frac{dC^k}{dt} - \sum T^i \frac{dD_i^k}{dt}\right).$$

Now

$$\frac{dC^k}{dt} = \sum_h \left(\frac{\partial \eta^k(w)}{\partial w^h}\right)_{w=y} \xi^h(y)$$

$$\frac{dD_i^k}{dt} = \frac{d}{dt} \frac{\partial(e^{t\xi}x)^k}{\partial x^i} = \sum_h \left(\frac{\partial \xi^k(w)}{\partial w^h}\right)_{w=y} D_i^h \qquad \text{(see 10.2).}$$

Hence

$$\sum_i T^i \frac{dD_i^k}{dt} = \sum_{ij} B_j^i C^j \frac{dD_i^k}{dt} = \sum_{ijh} B_j^i C^j D_i^h \left(\frac{\partial \xi^k(w)}{\partial w^h}\right)_{w=y}$$

$$= \sum_h C^h \left(\frac{\partial \xi^k(w)}{\partial w^h}\right)_{w=y} = \left(\sum_h \eta^h(w) \frac{\partial \xi^k(w)}{\partial w^h}\right)_{w=y}.$$

On substituting into (13.7a) we obtain finally

$$\frac{dT^i}{dt} = \sum B_k^i([\xi, \eta]y)^k. \tag{13.8}$$

We define a vector field $(t\xi)\circ\eta$ by the formula

$$((t\xi)\circ\eta)^i x = \sum_j B_j^i \eta^j(e^{t\xi}x). \tag{13.9}$$

The B_j^i (13.6) do not involve η. It follows that $(t\xi)\circ\eta$, regarded as a function of ξ, η defined over the totality of vector fields of class 1 over R, is linear in η. Formulas (13.7) and (13.8) now become (with x omitted)

$$\frac{d}{ds_0} T_{s,t} = (t\xi)\circ\eta, \tag{13.10}$$

$$\frac{d((t\xi)\circ\eta)}{dt} = (t\xi)\circ[\xi, \eta]. \tag{13.11}$$

14. Closed systems of class 1

Let $\xi_1, \cdots, \xi_r$ be vector fields of class 1 over $R \subset S_n$ and let L $(\xi_1, \cdots, \xi_r)$ be the totality of linear combinations of the ξ_i with real coefficients. If the commutator of every pair of elements in L is in L, we shall call L a *closed system* of class 1. (It is understood of course that a definite coordinate system in S_n has been chosen.)

Let ξ be an element in the closed system $L(\xi_1, \cdots, \xi_r)$ of class 1 and let $f_j^i(t) = (((t\xi)\circ\xi_j)x)^i$ where x is regarded as fixed. We assert that the real-valued functions $f_j^i(t)$ are analytic. For, $[\xi, \xi_j] = \sum C_{jk}\xi_k$ where the C's are constants. Then, omitting the index i for the moment, and using (3.11),

$$\frac{df_j}{dt} = (t\xi)\circ[\xi, \xi_j] = \sum C_{jk}((t\xi)\circ\xi_k) = \sum C_{jk}f_k .$$

Thus for each i, the f_j^i satisfy a system of linear homogeneous first order equations with constant coefficients, and assume definite values when $t = 0$. Hence they are convergent power series in t.

(14.1) *If ξ, η are in the closed system L $(\xi_1, \cdots, \xi_r)$ of class 1 then*

$$(t\xi)\circ\eta = \eta + t[\xi, \eta] + \frac{t^2}{2!}[\xi, [\xi, \eta]] + \cdots .$$

PROOF. We have seen that $((t\xi\circ\xi^j)x$ can be represented as a power series in t. We determine the coefficients. By (13.11)

$$\frac{d}{dt}(t\xi)\circ\xi_i = (t\xi)\circ[\xi, \xi_i]$$

$$\frac{d^2}{dt^2}(t\xi)\circ\xi_i = \frac{d}{dt}((t\xi)\circ[\xi, \xi_i]) = (t\xi)\circ[\xi, [\xi, \xi_i]]$$

. . . .

Setting $t = 0$ and using the obvious relation $(O\xi)\circ\xi_i = \xi_i$, we obtain the required expansion for $\eta = \xi_i$. It holds for arbitrary η since $(t\xi)\circ\zeta$ and $[\xi, \zeta]$ are linear in ζ.

COROLLARY (14.2). *If ξ, η belong to a closed system $L(\xi_1, \cdots, \xi_r)$ of class 1, then $(t\xi)\circ\eta$ belongs to L. As a linear combination of $\xi_1, \cdots, \xi_r$, $(t\xi)\circ\eta$ has coefficients which are convergent power series in t.*

Suppose that closure relations of the system $L(\xi_1, \cdots, \xi_r)$ in (14.1) are given by $[\xi_h, \xi_k] = \sum c^i_{hk}\xi_i$. Let C_h be the matrix (c^i_{hk}). On putting $\xi = \xi_h$ in (14.1) we find that

$$(t\xi_h)\circ\xi_k = \sum_j (e^{tC_h})^j_k \xi_j. \tag{14.3}$$

We shall make a useful application of this formula.

(14.4) Let c^k_{ij} be r^3 real numbers such that

$$c^k_{ij} = -c^k_{ji}, \qquad \sum_s (c^q_{is}c^s_{jk} + c^q_{js}c^s_{ki} + c^q_{ks}c^s_{ij}) = 0. \tag{14.5}$$

Let $C_m = (c^k_{mj})$. Then

$$e^{-tC_m}C_p e^{tC_m} = \sum_h (e^{tC_m})^h_p C_h. \tag{14.6}$$

PROOF. It follows from (14.5) that

$$C_hC_k - C_kC_h = \sum_i c^i_{hk} C_i.$$

Hence if we associate to each matrix C_h a vector field ξ_h by the formula $\xi^j_h = \sum c^j_{hi}x^i$, we find that $[\xi_h, \xi_k] = -\sum c^i_{hk}\xi_i$. Hence by (14.3)

$$((t\xi_m)\circ\xi_p)^i = \sum_{hj} (e^{-tC_m})^h_p c^i_{hj} x^j. \tag{14.7}$$

The left side equals (see 13.10)

$$\frac{d}{ds_0}(e^{-t\xi_m}e^{s\xi_p}e^{t\xi_m}x)^i = \frac{d}{ds_0}\sum_j (e^{-tC_m}e^{sC_p}e^{tC_m})^i_j x^j = \sum_j (e^{-tC_m}C_p e^{tC_m})^i_j x^j \tag{14.8}$$

from which (14.6) follows.

(14.9) Let ξ, η be vector fields in a closed system $L = L(\xi_1, \cdots, \xi_r)$ of class 1. Let X be the linear transformation $\zeta \to [\xi, \zeta]$ of L into itself. Then

$$e^{t\xi}e^{s\eta}x = e^{s\eta'}e^{t\xi}x \quad where \quad \eta' = e^{-tX}\eta. \tag{14.10}$$

PROOF. Evidently the transformation $T_{t,s}$ defined in §13 is additive in s: $T_{t,s}T_{t,w} = T_{t,s+w}$. Hence if we write $y = T_{t,s}x$ we have (cf. 12.3)

$$\frac{dy}{ds} = \frac{d}{d\sigma_0}T_{t,\sigma}y = ((t\xi)\circ\eta)y. \tag{14.11}$$

Now the expansion in (14.1) may be written $(t\xi)\circ\eta = e^{tX}\eta$.

The vector field $\eta' = e^{tX}\eta$ being in L by (14.2), is of class 1. Hence $e^{s\eta'}x$ is well defined and (14.11) implies that

$$e^{-t\xi}e^{s\eta}e^{t\xi}x = e^{s\eta'}x$$

which, on replacing t by $-t$ yields (14.10).

THEOREM (14.12). *Let L $(\xi_1, \cdots, \xi_r)$ be a closed system of class 1 over an open set R of a space S_n with coordinate system $x^1, \cdots, x^n$. Assume that the ξ_i are linearly independent over R. There exists a G_r with an analytic coordinate system $a^1, \cdots, a^r$ such that the function*

$$a \cdot x = e^{a^1\xi_1} \cdots e^{a^r\xi_r}x \tag{14.13}$$

defines an effective realization (G_r, S_n).

PROOF. Let $y = a \cdot x$. Then

$$\frac{\partial y}{\partial a^p} = \lim_{s\to 0} \frac{1}{s}\,(e^{a^1\xi_1}\cdots e^{(a^p+s)\xi_p}\cdots e^{a^r\xi_r}x - y).$$

Replace $e^{(a^p+s)\xi_p}$ by $e^{s\xi_p}\,e^{a^p\xi_p}$, then transfer $e^{s\xi_p}$ to the left step by step using (14.10). We obtain

$$\frac{\partial y}{\partial a^p} = \lim_{s\to 0} \frac{1}{s}\,(e^{s\zeta_p}y - y), \qquad \zeta_p = e^{-a^1X_1}\cdots e^{-a^pX_p}\xi_p$$

where X_i is the linear transformation $\xi \to [\xi_i, \xi]$. Thus

$$\frac{\partial(a\cdot x)}{\partial a^p} = \frac{d}{ds_0}\,e^{s\zeta_p}(a\cdot x) = \zeta_p(a\cdot x).$$

From (14.2) ζ_p is a linear combination of $\xi_1, \cdots, \xi_r$; its coefficients are power series in $a^1, \cdots, a^{p-1}$. Hence

$$\frac{\partial(a\cdot x)}{\partial a^p} = \sum_j \phi^j_p(a)\xi_j(a\cdot x) \tag{14.14}$$

where the ϕ_i are analytic. From (14.3) we see that $(\partial(a\cdot x)/\partial a^p)_{a^p=0} = \xi_p(x)$. Hence if we put $a = O$ in (14.14) we find that the matrix $(\phi^j_p(O))$ is the identity. Hence our theorem follows from (11.1).

For later reference we shall compute the ϕ_i more explicitly. Suppose $X_i\xi_j = [\xi_i, \xi_j] = \sum c^k_{ij}\xi_k$. Let $C_i = (c^k_{ij})$. Then $e^{tX_i}\xi_j = \sum(e^{tC_i})^k_j\xi_k$. Hence when $p > 1$,

$$\zeta_p = \sum \phi^k_p\xi_p = e^{-a^1X_1}\cdots e^{-a^{p-1}X_{p-1}}\xi_p = \sum(e^{-a^1C_1}\cdots e^{-a^{p-1}C_{p-1}})^k_p\,\xi_k, \tag{14.15}$$

$$\phi^k_p(a) = (e^{-a^1C_1}\cdots e^{-a^{p-1}C_{p-1}})^k_p. \tag{14.16}$$

When $p = 1$, $\phi^k_1(a) = \delta^k_1$.

15. The commutator

We now examine more closely the commutator function $[a, b]$ defined by (7.2), assuming the coordinate system of G_r to be canonical. We know that

$[a, b]$ is linear in b. We now show that $[a, b]$ is at least homogeneous in a. Let $\sigma = st$. Then if we use (7.2) and the fact that in canonical coordinates $ta = a^t$, we have

$$[ta, b] = \frac{d}{ds_0}((ta)^{-s}b(ta)^s) = \frac{d}{ds_0}(a^{-\sigma}ba^\sigma) = t\frac{d}{d\sigma_0}(a^{-\sigma}ba^\sigma) = t[a, b].$$

(15.1) *In analytic canonical coordinates*

$$[a, b] = \frac{d}{d\tau_0} b^{-t}a^{-t}b^t a^t, \qquad \tau = t^2.$$

PROOF. Using the homogeneity of $[a, b]$ we have (17.3)

$$a^{-t}b^s a^t = (-ta)(sb)(ta) = s\,(b + t[a, b] + \frac{t^2}{2!}[a,[a, b]] + \cdots] = sb' \quad \text{say}.$$

Using the expansion (4.1) we obtain

$$(15.2) \quad \begin{aligned}(b^{-s}a^{-t}b^s a^t)^i = (b^{-s}(a^{-t}b^s a^t))^i &= ((-sb)(sb'))^i = s(-b^i + b'^i) + s^2p_1 \\ &= st[a, b] + st^2p_2 + s^2p_3\end{aligned}$$

where p_1 and p_3 are power series in s, p_2 a power series in t. The coefficient of t^2 in (15.2) is zero. On evaluating $(b^{-s}a^{-t}b^s)\,a^t$ in a similar manner we see that, symmetrically, the coefficient of s^2 in (15.2) is zero. Hence the only term of second degree is $st\,[a, b]$. Therefore, on putting $s = t$ we obtain

$$\lim_{t\to 0}\frac{1}{t^2}(b^{-t}a^{-t}b^t a^t) = [a, b].$$

THEOREM (15.3). *Let G_r be a local group with analytic canonical coordinate system $a^1, \cdots, a^r$. Suppose there exists an effective realization of G_r defined by the function*

$$(15.4) \qquad a\cdot x = e^{a^1\xi_1 + \cdots + a^r\xi_r}x$$

where $\xi_1, \cdots, \xi_r$ are vector fields of class 1 over an open set R in a space S_n with coordinate system $x^1, \cdots, x^n$. Then the ξ_i are linearly independent over R and $L(\xi_1, \cdots, \xi_r)$ is closed. Moreover the linear correspondence $a \to \sum a^i\xi_i$ between G_r and L is such that if $a \to \xi$, $b \to \eta$, then $[a, b] \to [\xi, \eta]$.

PROOF. If the ξ_i are not linearly independent over R, we may suppose that $\xi_r \equiv c_1\xi_1 + \cdots + c_{r-1}\xi_{r-1}$ identically over an open subset R' of R. Then for every $a = (0, \cdots, 0, a^r)$ with $|a^r|$ sufficiently small we have $a\cdot x \equiv x$ over R' so that the realization could not be effective.—Suppose now that $a \to \xi = \sum a^i\xi_i$, $b \to \eta = \sum b^i\xi_i$. Then $e^{t\xi}x = (ta)\cdot x = a^t\cdot x$, $e^{t\eta}x = b^t\cdot x$. Hence

$$T_{t,s}x = e^{-t\xi}e^{s\eta}e^{t\xi}x = a^{-t}b^s a^t\cdot x,$$

or, by (7.3) and the homogeneity of $a^{-1}ba$,

$$T_{t,s}x = (sb')\cdot x,\ b' = b + t[a, b] + \cdots.$$

On taking the derivative of both sides of (15.4) with respect to s at $s = 0$ we have (see 13.10, 15.4)

$$(t\xi)\circ\eta = \sum \left(\frac{\partial(b\cdot x)}{\partial b^i}\right)_{b=0} b'^i = \sum b'^i \xi_i .$$

Now take the derivative with respect to t at $t = 0$ using (13.11). We obtain

$$[\xi, \eta] = \sum [a, b]^i \xi_i .$$

Hence L is closed and $[a, b] \to [\xi, \eta]$.

(15.5) *In an analytic canonical coordinate system,* $[a, b]$ *satisfies the identities*

(15.6) $\quad [a, b] = -[b, a], [a, [b, c]] + [b, [c, a]] + [c, [a, b]] = O.$

PROOF. The given G_r always possesses a realization (G_r, S_n) satisfying the hypotheses of (15.3) (the regular realization for example, by virtue of (12.9)). The identities in question then follow by (15.3) from the corresponding identities (13.1) and (13.2).

COROLLARY (15.7). $[a, b]$ *is linear in both a and b.*

(15.8) Let G_r, G'_s be local groups with analytic canonical coordinates a^i and a'^i respectively. Suppose there exists a local homomorphism τ: $G_r \to G'_s$. Then $s \leqq r$. Moreover, if $\tau a = a'$, $\tau b = b'$, then $[a, b] = [a', b']$.

PROOF. By (5.4) τ induces a linear mapping of the space G_r onto G'_s. Hence s can not exceed r. Moreover, since $\tau(ta) = ta'$, we have

$$\tau\left(\frac{1}{t^2}(-tb)(-ta)(tb)(ta)\right) = \frac{1}{t^2}((-tb')(-ta')(tb')(ta')).$$

Hence by (15.1), $[a, b] = [a', b']$.

16. Lie groups and algebras

We shall say that a local group G is an *r-parameter Lie group* if it is locally isomorphic to a G_r with left-differentiable coordinates satisfying a right Lipschitz condition. To make it clear that this definition admits no ambiguity concerning the number r, we remark that if G is locally isomorphic to G_s—a local group of the same type as G_r—then $r = s$. This is an immediate consequence of (15.8).

An r-parameter (real) *Lie algebra* L_r is an r-dimensional vector space $\{a, b, \cdots\}$ over the field of real numbers, with a bilinear non-associative "product" $[a, b]$ satisfying (15.6).

Suppose G_r has an analytic canonical coordinate system $a^1, \cdots, a^r$. Then it follows from (15.5, 15.7) that if we use the commutator $[a, b]$ for product, G_r is converted into an r-parameter Lie algebra. We denote this algebra by $L(G_r)$.

Let us now regard two Lie algebras as equivalent if they are isomorphic, and two Lie groups if they are locally isomorphic. Then the totality of r-parameter Lie groups falls into a system Γ_r of equivalence classes and the algebras fall into equivalence classes $\mathfrak{L}_r$.

To each element γ_r of Γ_r there is associated a unique element $\mathfrak{l}(\gamma_r)$ of $\mathfrak{L}_r$ in the following way. By (12.4), γ_r contains a G_r with analytic canonical coordinates $a^1, \cdots, a^r$. We take for $\mathfrak{l}(\gamma_r)$ that equivalence class of Lie algebras which con-

tains $L(G_r)$. It is of course necessary to remark that if G'_r is a local group with analytic canonical coordinates $a'^1, \cdots, a'^r$ and if G'_r is locally isomorphic to G_r then $L(G'_r)$ is locally isomorphic to $L(G_r)$. This follows from (15.8).

(16.1) *The correspondence $\gamma_r \to \mathfrak{L}(\gamma_r)$ is one-one between Γ_r and $\mathfrak{L}_r$.*

PROOF. We note first that the correspondence $\gamma_r \to \mathfrak{l}(\gamma_r)$ has a single-valued inverse. For, isomorphism between $L(G_r)$ and $L(G'_r)$ implies local isomorphism between G_r and G'_r. This is a consequence of (12.8).

It remains to be shown that for every element $\mathfrak{l}_r$ of $\mathfrak{L}_r$ there exists an element γ_r of Γ_r such that $\mathfrak{l}(\gamma_r) = \mathfrak{l}_r$. For this it is sufficient to show that if L_r is a given r-parameter Lie algebra, there exists a G_r such that $L(G_r) = L_r$.

We make a preliminary remark. A second method of associating a Lie algebra with a given G_r with analytic canonical coordinates $a^1, \cdots, a^r$ is the following. Let (G_r, S_n) be an effective realization of G_r such that relative to some coordinate system $x^1, \cdots, x^n$ in S_n the functions $(a\cdot x)^i$ have continuous second derivatives with respect to $a^1, \cdots, a^r$. Such a realization always exists,—the regular realization, for example. The vector fields

$$\xi_i = \left(\frac{\partial(a\cdot x)}{\partial h^i}\right)_{a=0}$$

are of class 1 and

$$a\cdot x = a^{a^1\xi_1+\cdots+a^r\xi_r}x. \tag{16.2}$$

(The proof of (16.2) is contained in (12.3) with ψ_i replaced by ξ_i and $a^t b$ by $(ta)\cdot x$.) By (15.3), $\xi_1, \cdots, \xi_r$ are linearly independent and $L(\xi_1, \cdots, \xi_r)$ is a closed system. Hence $L(\xi_1, \cdots, \xi_r)$ may be regarded as an n-dimensional n-parameter Lie algebra $L_\xi(G_r)$ associated with G_r. By (15.3) *$L_\xi(G_r)$ is isomorphic to $L(G_r)$*

From these remarks it follows by (14.12) that in order to complete the proof of (16.1), it is sufficient to show that in a space S_r with coordinate system $x^1, \cdots, x^r$ there exist vector fields $\xi_1, \cdots, \xi_r$ linearly independent and of class 1 over some open set R and such that, as an r-parameter Lie algebra, $L(\xi_1, \cdots, \xi_r)$ is isomorphic to the given Lie algebra L_r, the ξ's being such that L is closed.

Let $a_1, \cdots, a_r$ be a basis for L_r, and let $[a_j, a_k] = \sum c^i_{jk}a_i$. It will be sufficient to find ξ's such that $[\xi_j, \xi_k] = \sum c^i_{jk}\xi_i$, that is such that

$$[\xi_j, \xi_k]^h = \sum\left(\xi^i_j\frac{\partial\xi^h_k}{\partial x^i} - \xi^i_k\frac{\partial\xi^h_j}{\partial x^i}\right) = \sum_i c^i_{jk}\xi^h_i. \tag{16.3}$$

To show that such ξ's exist, it is sufficient to show that there exist vector fields $\phi_1, \cdots, \phi_r$ in S_r which are analytic in the neighborhood of $x = 0$, are such that the matrix $(\phi^i_j(O))$ is non-singular, and which satisfy the equations

$$\frac{\partial\phi^i_p}{\partial x^q} - \frac{\partial\phi^i_q}{\partial x^p} = \sum_{jk} c^i_{jk}\phi^j_p\phi^k_q. \tag{16.4}$$

For, if we take for $(\xi^i_j(x))$ the inverse of the matrix $(\phi^i_j(x))$ we obtain vector fields ξ_j which satisfy (16.3) (the verification of this is straightforward). Since the matrix $(\xi^i_j(O))$ is non-singular, the ξ's will be analytic and linearly independent in some neighborhood of $x = 0$

We shall show that the functions ϕ_j^i given by (14.16) (with a replaced by x) have the required properties. Evidently these functions are analytic and the matrix $(\phi_j^i(O))$ is the identity. To verify (16.4), suppose $p < q$. Then $\partial\phi_p^i/\partial x^q = 0$. Hence, denoting the left side of (16.4) by Φ and writing $\gamma_i = x^i C_i$, we have

$$\Phi = -\frac{\partial \phi_q^i}{\partial x^p} = (e^{-\gamma_1} \cdots C_p \cdots e^{-\gamma_{q-1}})_q^i. \tag{16.5}$$

Now on account of (13.1, 13.2) the numbers c_{jk}^i satisfy (14.5). Hence we may use (14.6) writing it in the form

$$e^{-\gamma_m} C_p = \sum_j (e^{-\gamma_m})_p^j (C_j e^{-\gamma_m}).$$

We use this formula in (16.5) to shift C_p to the left, step by step, giving m successively the values $p - 1, \cdots, 0$. We obtain

$$\begin{aligned}\Phi &= \sum_j (e^{-\gamma_1} \cdots e^{-\gamma_{p-1}})_p^j (C_j e^{-\gamma_1} \cdots e^{-\gamma_{q-1}})_q^i \\ &= \sum_j \phi_p^j (C_j e^{-\gamma_1} \cdots e^{-\gamma_{q-1}})_q^i = \sum_{jk} \phi_p^j c_{jk}^i \phi_q^k.\end{aligned}$$

The proof in the case $p > q$ is the same. When $p = q$ both sides of (16.4) vanish identically (the right side because $c_{jk}^i = -c_{kj}^i$).

17. Local subgroups of Lie groups

Let us regard a local group consisting of a single element as a "G_0 with analytic canonical coordinates". Then it follows directly from (5.7) and the definition of Lie group that *every local subgroup of a Lie group is a Lie group.*

Let H be a local subgroup of a G_r with analytic canonical coordinates. By (15.7) H may be regarded as imbedded in a flat subspace H^0 of G_r. Regarded as a subset of the Lie algebra $L(G_r)$, H^0 is evidently a subalgebra (this follows from (15.1)) and this subalgebra is precisely $L(H)$.

(17.1) *If H is a flat subspace of G_r with analytic canonical coordinates $a^1, \cdots, a^r$ and if H, regarded as a subset of $L(G_r)$ is a subalgebra L^0 of* $L(G_r)$, *then H is a local subgroup of G_r.*

Proof. Since canonical coordinates will remain canonical after undergoing a linear homogeneous transformation, we may suppose that H is the $(a^1, \cdots, a^p)$-coordinate subspace of G_r. Now let ψ_i be the vector fields defined by (3.1). We have (12.9)

$$ab = e^{\Sigma a^i \psi_i} b \ (a = (a^1, \cdots, a^r)). \tag{17.2}$$

By (15.3), $L(\psi_1, \cdots, \psi_r)$ is a Lie algebra isomorphic to $L(G_r)$. The isomorphism is given by: $a \rightarrow \sum a^i \psi_i$. Hence the subgroup of $L(\psi_1, \cdots, \psi_r)$ which corresponds to H is precisely $L(\psi_1, \cdots, \psi_p)$. Since this subsystem is closed, it follows from (14.12) that the function

$$a \cdot b = e^{a^1\psi_1} \cdots e^{a^p\psi_p} b \quad \text{where} \quad a = (a^1, \cdots, a^p, 0, \cdots, 0)$$

defines a realization of a p-parameter group G_p (which, as a space, is identical with H). Let a_i denote the projection on the a^i-axis of an arbitrary element a of $G_r : a_i = (0, \cdots, a^i, \cdots, 0)$. It follows from (17.2) that when $a \in G_p = H$,

$$(17.3) \qquad a \cdot b = a_1 \cdots a_p b, \qquad \text{hence} \quad a \cdot O = a_1 \cdots a_p .$$

Now let a, c be arbitrary elements of H near O, and let $d = ca$ where ca is the group product *defined in the group* G_p. From (17.3) we have

$$d_1 \cdots d_p = d \cdot O = c \cdot (a \cdot O) = c_1 \cdots c_p a_1 \cdots a_p .$$

We draw the following conclusion: consider elements of G_r of the form $a_1 \cdots a_p$ where a_i is an arbitrary element on the a^i-axis and $|a_i| < \delta$, δ small. The totality of these elements *is a local subgroup* of G_r. By (5.7) this local subgroup can be regarded as part of a linear subspace H' of G. Evidently H' contains the a^1-, $\cdots$, a^p-axes. Hence H' contains H. It is easy to see that the dimension of H' can not exceed that of H. Hence $H' = H$.

(17.4) Suppose H is a local subgroup of G_r so that $L(H)$ is a subalgebra of $L(G_r)$. If H is normal (i.e. if, near O, $b \in H$ implies $aba^{-1} \in H$) then so is $L(H)$ (i.e. $b \in H$ implies $[a, b] \in H$) and conversely. The first statement follows from (15.1), the converse from (7.2).

APPENDIX

Let R be a spherical region in a euclidean space E_n, B its boundary. Let σ be a continuous mapping of $\bar{R} = R + B$ in E_n such that for each point b in B, the vector $\overrightarrow{b, \sigma b}$, if it is not of length 0, makes an acute angle with the outward normal to B at b. Then $R \subset \sigma R$.

The proof can be based on the concept of the degree of a mapping (see for example: Lefschetz, *Algebraic Topology*). Suppose that, contrary to the conclusion, there exists a point x in R not covered by σR. Using the boundary conditions for σ, it is easy to see that, without disturbing the relation $x \notin \sigma R$, σ can be made to pass continuously into a mapping σ' which transforms $\bar{R}$ into itself and in particular leaves every point on B invariant. Since x is not covered by $\sigma' R$, the degree of σ', i.e. the algebraic number of times that $\sigma' R$ covers R, is zero. But by standard topological methods it can be proved that the degree of σ' equals the degree of the mapping $B \to B$ induced by σ', namely $+1$. This contradiction establishes the theorem.

COLUMBIA UNIVERSITY.

REFERENCES

1. GARRETT BIRKHOFF, *Analytical groups*, Trans. Am. Math. Soc. 43 (1938) 61–101.
2. E. CARTAN, *La théorie des groupes finis et continus et l'analyse situs*, Mem. des Sciences (1930) Fasc. XLII.
3. E. R. VAN KAMPEN, *Remarks on systems of ordinary differential equations*, Am. J. 59 (1937) 144–152.
4. B. L. VAN DER WAERDEN, *Göttingen lectures on continuous groups* (mimeographed).
5. J. H. C. WHITEHEAD, *A note on Maurer's equations*, J. Lond. Math. Soc. 7 (1932) 223–227.

ANNALS OF MATHEMATICS
Vol. 44, No. 3, July, 1943

COMBINATIONS OF CLOSURE RELATIONS

BY OYSTEIN ORE

(Received October 27, 1942)

The present paper contains a study of the system of all possible closure relations definable over a given set. It has already been shown by Garrett Birkhoff [1] that this system may be considered to be a structure so that the interrelations of the various closure relations indicate properties of this structure. One finds that the structure is a complete dual point structure satisfying the so-called Birkhoff condition. It is found furthermore that to every closure relation there exists a unique maximal closure relation relatively prime to it. From this result one obtains a condition for two closure relations to be complementary. The implications of the Dedekind law are studied briefly. The application of one closure relation after another is called the product of the two and a necessary and sufficient condition for the product of the two closure relations to be a closure relation is given.

It is shown that the automorphisms of the structure of closure relations are all induced by the one-to-one transformations of the basic set. By means of a general theorem on the homomorphisms of point structures all complete homomorphisms are determined. To conclude there are a few remarks about closure relations in which every point is closed or also determined by its closure.

In another paper some similar investigations will be carried through for additive closure relations.

1. Basic properties of closure operations

In the following S shall denote a fixed set. A correspondence Γ which associates with every subset A of S some unique subset $\bar{A}$ of S

$$A \rightarrow \bar{A} = \Gamma(A)$$

shall be called a *closure relation* or *closure operation* when it has the three properties:

Closure:

$$\textit{Idempotent: } \bar{\bar{A}} = \bar{A}$$

$$\textit{Increasing: } \bar{A} \supseteq A$$

$$\textit{Order preserving: } A_1 \supseteq A_2 \text{ implies } \bar{A}_1 \supseteq \bar{A}_2 .$$

The set $\bar{A}$ is called the *closure of* A *under* Γ and any such image set $\bar{A}$ is *closed under* Γ. The closure of the whole set is S and by an additional definition one prescribes that the void set 0 is closed.

Let us mention a few simple properties which follow directly from the definition of a closure relation:

A set is closed if and only if $A = \bar{A}$.

The closure $\bar{A}$ of a set A is the smallest closed set containing A.

For the closures of sums and intersections of the sets in some family $\mathfrak{F}(A_i)$ of subsets of S one finds

$$(1) \qquad \sum_i \bar{A}_i = \overline{\sum_i A_i} \supseteqq \sum \bar{A}_i$$

and

$$(2) \qquad \overline{\prod \bar{A}_i} = \prod \bar{A}_i \supseteqq \overline{\prod A_i}.$$

From (2) one concludes further:

The intersection of any family of closed sets is closed.

An *intersection ring of sets* in S is a family of subsets which includes the intersection of any two of its sets and a *complete intersection ring* contains the intersection of any subfamily of its sets. An intersection ring of sets *over* S is defined to contain the sets 0 and S. This terminology permits us to express the following basic result (see *E. H. Moore* [1]):

THEOREM 1. *In a closure relation the closed sets form a complete intersection ring of sets over S. Conversely any closure relation is obtained from a complete intersection ring $\mathfrak{R}(\bar{A})$ over S by associating with each set A the smallest set $\bar{A}$ in $\mathfrak{R}(\bar{A})$ containing A.*

A set S in which a closure relation has been defined is also called a *topological space* and instead of a closure relation we may use the term *topology*. Any family of sets in S defines a topology through the complete intersection ring of sets over S which it generates. For a given closure relation any family of sets from which all closed sets can be obtained by forming set intersections is called a *basis for the closed sets*. As usual the complement of a closed set is an *open set*. The open sets form a *complete sum ring of sets over* S, and any family of open sets from which all other open sets can be obtained by summation is called a *basis for the open sets.*

2. The structure of closure relations

The system of all closure relations in a set becomes a partially ordered set by the following definition of an inclusion: Let Γ_1 and Γ_2 be two closure relations and $\Gamma_1(A)$ and $\Gamma_2(A)$ the closures of a set A in the two topologies. We shall say that Γ_1 *contains* Γ_2 and write $\Gamma_1 \supseteqq \Gamma_2$ when one has

$$\Gamma_1(A) \supseteqq \Gamma_2(A)$$

for every subset A of S. From this definition one obtains:

THEOREM 2. *The necessary and sufficient condition for a closure relation Γ_1 to contain another closure relation Γ_2 is that any set closed under Γ_1 is closed under Γ_2.*

PROOF: When Γ_1 contains Γ_2 one has for any subset A of S

$$\Gamma_1(\Gamma_1(A)) = \Gamma_1(A) \supseteqq \Gamma_2(\Gamma_1(A)) \supseteqq \Gamma_1(A)$$

so that

$$\Gamma_1(A) = \Gamma_2(\Gamma_1(A))$$

and $\Gamma_1(A)$ is closed under Γ_2. This may also be stated that the complete intersection ring $\mathfrak{R}(\Gamma_1)$ of closed sets under Γ_1 is contained in the ring $\mathfrak{R}(\Gamma_2)$ of closed sets under Γ_2. Conversely since for any closure relation Γ the closure $\Gamma(A)$ is the least closed set containing A one must have

$$\Gamma_1(A) \supseteqq \Gamma_2(A)$$

whenever

$$\mathfrak{R}(\Gamma_2) \supseteqq \mathfrak{R}(\Gamma_1).$$

We prove next a result due to *Garrett Birkhoff* [1]:

THEOREM 3. *The closure relations over a set form a complete structure.*

PROOF: One sees that in the partially ordered set of closure relations the union $V\Gamma_i$ of a set of closure relations Γ_i is the closure relation defined by the sets common to all the intersection rings defined by the various Γ_i. The intersection $\wedge \Gamma_i$ of the same closure relations is the closure relation defined by the complete intersection ring generated by all sets closed under any Γ_i.

The system of all complete intersection rings of sets over S form a structure in which the cross cut $\mathfrak{R}_1 \cap \mathfrak{R}_2$ of two rings $\mathfrak{R}_1$ and $\mathfrak{R}_2$ consists of their common sets and the union $\mathfrak{R}_1 \cup \mathfrak{R}_2$ is the ring generated by the sets in $\mathfrak{R}_1$ and $\mathfrak{R}_2$. From the proof of theorem 3 one concludes:

THEOREM 4. *The structure of all complete intersection rings over a set S is dually isomorphic to the set of all topologies in S.*

In connection with the definition of the cross-cut of several closure relations one should mention the following result:

THEOREM 5. *Any intersection*

$$D = \prod_i \Gamma_i(A_i)$$

of closed sets $\Gamma_i(A_i)$ from a system $\{\Gamma_i\}$ of closure relations can be represented in the form

$$D = \prod_i \Gamma_i(D).$$

PROOF: Since

$$\Gamma_i(A_i) \supset D$$

one obtains

$$\Gamma_i(\Gamma_i(A_i)) = \Gamma_i(A_i) \supset \Gamma_i(D)$$

consequently

$$\prod_i \Gamma_i(A_i) = D \supset \prod_i \Gamma_i(D) \supset D$$

so that the theorem follows. Theorem 5 can also be stated in the form:

THEOREM 6. *The intersection* $\bigwedge_i \Gamma_i$ *of a system of closure relations is the closure relation in which the closure of a set is the intersection of its closures in the various relations* Γ_i.

The structure of closure relations has a universal element V containing all others. This *universal closure relation* is defined by the property that

$$V(A) = S$$

for every set $A \neq 0$. Similarly the structure has a zero element contained in all others. This is the *identical closure relation* I for which

$$\mathrm{I}(A) = A$$

for every subset A of S.

The structure of closure relations contains maximal elements M contained in no other closure relations except V. Such a *maximal* closure relation must have an intersection ring of closed sets consisting of the trivial sets O and S and a single other closed set A. We shall call A the *principal set* of M and write $\mathrm{M} = \mathrm{M}_A$. According to the definition of inclusion for closure relations one sees that a closure relation Γ is contained in the maximal closure relations $M_{\bar{A}}$ where $\bar{A}$ is a set closed under Γ. Obviously one has:

THEOREM 7. *Any closure relation is the cross-cut of the maximal closure relations in which it is contained.*

The structure of closure relations also has minimal elements containing I and no other closure relation. Among the closed sets in a minimal closure relation N one cannot have all maximal sets $S - a$, where a is some element, because every subset is the intersection of such maximal sets and one would have $\mathrm{N} = \mathrm{I}$. But on the other hand if a closure is to be minimal it cannot omit from its closed sets more than one such maximal set. Thus we see that the minimal closure relations are those in which every set is closed except a single maximal set $S - a$. This can also be stated in the form:

THEOREM 8. *The minimal closure relations* N_a *are those in which every set except the point a is open.*

It should be noted that the dual of theorem 7 does not hold: Not every closure relation is the union of the minimal closure relations which it contains. To see this we observe that every minimal closure relation has among its closed sets all those which omit two or more elements of S and it is not difficult to verify that only closure relations with this property can be written as the union of the minimal closure relations they contain.

A structure is called a *point structure* if every one of its elements is the union of the points or minimal elements it contains. If every element is the intersection of the maximal elements in which it is contained the structure may be called a *dual point structure*. Thus we have:

THEOREM 9. *The structure of topologies is a complete dual point structure.*

According to theorem 4 the structure of complete intersection rings is a complete point structure.

A closure relation Γ may contain certain maximal sets $S - d$ among its closed sets, hence also certain points d which are open sets i.e. *isolated points.* We shall denote by

$$O_\Gamma = \{d\}$$

the set of all isolated points for the closure relation Γ. This concept enters into the following considerations.

Two closure relations Γ_1 and Γ_2 are said to be *relatively prime* when $\Gamma_1 \cap \Gamma_2 = I$. This means that the intersection of the closed sets in the two closure relations constitute all subsets of S. Dually when $\Gamma_1 \cup \Gamma_2 = V$ the corresponding intersection rings of closed sets have no sets in common except O and S. We shall now characterize the closure relations which are relatively prime to a given relation Γ. Let $\Gamma \cap \Gamma^* = I$ so that every set is the intersection of two sets closed under Γ and Γ^* respectively. Since a maximal set $S - a$ cannot be obtained as the intersection of two larger sets, one concludes that any such set must be closed either under Γ or under Γ^*. But conversely when this is the case any set in the intersection of sets closed under Γ and Γ^* so that we conclude:

THEOREM 10. *Two closure relations Γ and Γ^* are relatively prime if and only if every point is an isolated point in at least one of these topologies, hence if*

$$O_\Gamma + O_{\Gamma^*} = S.$$

From this analysis also follows that among the various closures relatively prime to a given closure Γ there exists a unique maximal one. This is the closure Γ' whose closed sets are generated by those maximal sets $S - c$ which are not closed under Γ. Thus we can state:

THEOREM 11. *To any closure relation Γ there exists a unique maximal closure relation Γ' relatively prime to Γ. This closure relation is defined by the property that the points in*

$$O_{\Gamma'} = S - O_\Gamma$$

form a basis for its open sets.

From these results it is easy to derive the condition for a closure relation to have a complement in the structure of all closure relations. If Γ^* is to be a complement of the closure relation Γ one must have

$$\Gamma \cup \Gamma^* = V, \qquad \Gamma \cap \Gamma^* = I.$$

The last of these conditions states that Γ^* shall be relatively prime to Γ, hence one sees that if a complement exists the maximal relatively prime relation Γ' defined by theorem 11 must be such a complement, in fact maximal among all complements. Since Γ and Γ' shall have intersection rings of closed sets without common sets except 0 and S it follows that Γ can have no closed sets of the form $S - C'$ where C' is a subset of the set $O_{\Gamma'}$ of isolated points in Γ'. This is equivalent to saying that no open set under Γ can be a subset of $O_{\Gamma'}$ so that we have:

Theorem 12. *There exists a complement to a closure relation* Γ *if and only if no set open under* Γ *is a subset of the set* $S - O_\Gamma$ *of all non-isolated points under* Γ. *If a complement exists there exists a unique maximal one containing all others, namely the maximal closure relation* Γ' *relatively prime to* Γ.

Using the ordinary structure terminology we shall say that a closure relation Γ_1 is *prime over* another Γ_2 or Γ_2 is *prime under* Γ_1 when $\Gamma_1 \supset \Gamma_2$ and there exists no closure relation between Γ_1 and Γ_2. This implies of course that there can be no complete intersection ring of sets between two such rings $\mathfrak{R}_1$ and $\mathfrak{R}_2$ consisting of the closed sets in Γ_1 and Γ_2 respectively. Consequently it must be possible to obtain $\mathfrak{R}_2$ from $\mathfrak{R}_1$ through the adjunction of a single set A_2 to $\mathfrak{R}_1$. Let us form the intersection $A_2' = A_1A_2$ of A_2 with some set A_1 in $\mathfrak{R}_1$. This set A_2' must belong to $\mathfrak{R}_1$ because otherwise one could have adjoined A_2' instead of A_2 to $\mathfrak{R}_1$ and one would have obtained a complete intersection ring between $\mathfrak{R}_1$ and $\mathfrak{R}_2$ against assumption. Thus a closure relation Γ_2 is prime under a closure relation Γ_1 if the ring $\mathfrak{R}_2$ corresponding to Γ_2 contains some set A_2 which is not the intersection of other sets in $\mathfrak{R}_2$ and $\mathfrak{R}_1$ is obtained by omitting A_2 from $\mathfrak{R}_2$.

An important property of the structure of closure relations is expressed in the theorem:

Theorem 13. *Let* $\Gamma_1 \supset \Gamma_2$ *be two closure relations, one prime over the other, and let* Γ *be an arbitrary closure relation. Then the closure relation* $\Gamma_1 \cup \Gamma$ *is equal to or prime over* $\Gamma_2 \cup \Gamma$.

Proof: Let us denote by $\mathfrak{R}_1$, $\mathfrak{R}_2$ and $\mathfrak{R}$ respectively the intersection rings of closed sets corresponding to Γ_1, Γ_2 and Γ. According to the preceding remarks $\mathfrak{R}_2$ contains a single set A_2 not in $\mathfrak{R}_1$. The ring corresponding to $\Gamma_1 \cup \Gamma$ consists of the common sets in $\mathfrak{R}_1$ and $\mathfrak{R}$ and the ring corresponding to $\Gamma_2 \cup \Gamma$ of the sets common to $\mathfrak{R}_2$ and $\mathfrak{R}$. If these two rings are not equal they can at most differ by the set A_2 and theorem 13 follows.

An immediate consequence of this theorem is:

If the two closure relations Γ_1 *and* Γ_2 *are both prime over* $\Gamma_1 \cap \Gamma_2$ *then* $\Gamma_1 \cup \Gamma_2$ *is prime over both* Γ_1 *and* Γ_2.

From this so-called *Birkhoff property* follows as in a general structure that the following *chain theorem* holds:

If two closure relations $\Gamma \supset \Gamma'$ *are joined by a finite chain of closure relations, each prime over the next, then any two such chains between* C *and* C' *have the same length.*

For a finite set S the number of closure relations in a complete chain from I to V is found to be 2^n where n is the number of elements in S.

3. Analysis of the Dedekind law

We shall turn next to the study of the so called *Dedekind law*

$$\text{(3)} \qquad \mathrm{A} \cap (\mathrm{B} \cup \Gamma) = \mathrm{B} \cup (\mathrm{A} \cap \Gamma), \qquad \mathrm{A} \supset \mathrm{B}$$

in the structure of all closure relations in a set. It is not difficult to verify by

examples that this relation does not hold in general for closure relations. Let us establish what the condition (3) implies for the closed sets under the three closure relations in question. The sets which are closed under $\mathrm{A} \cap \Gamma$ are the intersections $A_1 \cdot C_1$ where A_1 is closed under A and C_1 under Γ. Therefore the sets closed under $\mathrm{B} \cup (\mathrm{A} \cap \Gamma)$ are those which are simultaneously closed under B and $\mathrm{A} \cap \Gamma$, i.e. the sets B_1 closed under B which are representable in the form

$$B_1 = A_1 \cdot C_1 . \tag{4}$$

The sets closed under $\mathrm{B} \cup \Gamma$ are those which are closed simultaneously under B and Γ

$$D_2 = B_2 = C_2$$

hence the sets closed under $\mathrm{A} \cap (\mathrm{B} \cup \Gamma)$ are the intersections $A_2 \cdot D_2$ where A_2 is closed under A. If the relation (3) is to hold there must exist for every set of the form (4) some representation

$$B_1 = A_1 \cdot C_1 = A_2 \cdot D_2 .$$

Since one can obviously assume that

$$A_2 = \mathrm{A}(B_1), \qquad D_2 = (\mathrm{B} \cup \Gamma)(B_1)$$

we can say:

THEOREM 14. *The necessary and sufficient condition for the Dedekind relation* (3) *to hold for the three closure relations* A, B *and* Γ *is that any set B_1 closed under* B *which is the intersection*

$$B_1 = A_1 \cdot C_1$$

of a set A_1 closed under A *and a set C_1 closed under* Γ *shall also be representable in the form*

$$B_1 = A_1 \cdot C_1 = \mathrm{A}(B_1) \cdot (\mathrm{B} \cup \Gamma)(B_1).$$

We shall use this theorem to obtain conditions for the Dedekind relation to hold identically in one of the three closure relations when the other two are fixed. Let us assume first that B and Γ are given closure relations and we wish to determine which properties they must have in order that (3) shall be fulfilled for all closure relations A containing B. In order to establish a necessary condition we shall take $\mathrm{A} = \mathrm{M}_{B_i}$ as a maximal closure relation. Here A contains B if and only if the principal set B_i in A is a set closed under B. From theorem 14 one concludes that if an intersection $B_i \cdot C_1$ is closed under B one must have

$$B_1 = B_i \cdot C_1 = \mathrm{A}(B_1) \cdot (\mathrm{B} \cup \Gamma)(B_1).$$

There are only three possibilities, namely 0, S and B_i for the set $\mathrm{A}(B_1)$ and the first of these is excluded when we make the trivial assumption that B_1 is not void. When $\mathrm{A}(B_1) = S$ then B_1 becomes closed under Γ. Finally when $\mathrm{A}(B_1) = B_i$ one must have

$$B_1 = B_i \cdot C_1 = B_i \cdot (\mathrm{B} \cup \Gamma)(B_1) \tag{5}$$

and this condition is seen to be satisfied in all three cases. But conversely when this condition (5) is fulfilled for all sets B_i the requirements of theorem 14 are also satisfied so that we can state:

THEOREM 15. *The necessary and sufficient condition for the Dedekind law* (3) *to hold for all closure relations A containing B is that whenever the intersection $B_1 \cdot C_1$ of a set B_1 closed under B and C_1 closed under Γ is itself closed under B there shall also exist a representation*

$$B_1 \cdot C_1 = B_1 \cdot (\mathrm{B} \cup \Gamma)(B_1 \cdot C_1).$$

It may be observed further that if the relation (3) holds for all A one also has

$$\mathrm{A} = \mathrm{B} \cup (\mathrm{A} \cap \Gamma) \tag{6}$$

for every A such that

$$\mathrm{B} \cup \Gamma \supset \mathrm{A} \supset \mathrm{B}. \tag{7}$$

Conversely when (6) holds for all A satisfying (7) the relation (3) is fulfilled for all A.

Let us now assume in (3) that A and Γ are fixed while B represents an arbitrary closure relation contained in A. A particular closure relation B contained in A shall be constructed in the following way. We denote by $B_1 = A_1 \cdot C_1$ the intersection of a set A_1 closed under A and a set C_1 closed under Γ and adjoin B_1 to the sets closed under A to obtain the intersection ring of sets closed under B. Thus the closed sets under B are either closed under A or they have the form

$$B_i = A_i \cdot A_1 \cdot C_1 = A'_i \cdot C_1$$

where A'_i is an arbitrary closed set under A which is contained in A_1. Since we can assume A $\neq$ B the set B_1 is not closed under A. Furthermore the set B_1 is obviously the largest of all sets closed under B but not under A. From theorem 14 one concludes that if (3) shall hold one must have

$$B_1 = A_1 \cdot C_1 = \mathrm{A}(B_1) \cdot (\mathrm{B} \cup \Gamma)(B_1).$$

The set B $\cup$ $\Gamma(B_1)$ cannot be closed under A because then B_1 would become closed under A against our assumption. From

$$(\mathrm{B} \cup \Gamma)(B_1) = A'_i \cdot C_1 \subseteq B_1$$

one concludes therefore that

$$(\mathrm{B} \cup \Gamma)(B_1) = B_1$$

so that the intersection $A_1 \cdot C_1$ becomes closed under Γ. Thus we have the necessary condition that every intersection $A_1 \cdot C_1$ which is not closed under A shall be closed under Γ and this condition is seen to be sufficient according to theorem 14 so that we can state:

THEOREM 16. *The necessary and sufficient condition for the Dedekind law* (3) *to hold for a fixed pair of closure relations* A *and* Γ *and for any closure relation* B

contained in Α *is that any intersection* $A_1 \cdot C_1$ *of a closed set* A_1 *under* Α *and* C_1 *under* Γ *be closed either under* Α *or* Γ.

Since this condition is symmetric in Α and Γ we have also obtained the further result:

THEOREM 17. *When*

$$\mathrm{A} \cap (\mathrm{B} \cup \Gamma) = \mathrm{B} \cup (\mathrm{A} \cap \Gamma)$$

for every closure relation B *contained in* Α *one also has*

$$\Gamma \cap (\mathrm{B} \cup \mathrm{A}) = \mathrm{B} \cup (\Gamma \cap \mathrm{A})$$

for every B *contained in* Γ.

Another observation to be made in this connection is that when (3) holds for all B contained in Α one has

$$\mathrm{B} = \mathrm{A} \cap (\mathrm{B} \cup \Gamma)$$

for every B such that

$$\mathrm{A} \supset \mathrm{B} \supset \mathrm{A} \cap \Gamma$$

and the converse is also true.

In the third and final case of the Dedekind law we shall investigate when the condition (3) holds for all closure relations Γ for a fixed pair of closure relations $\mathrm{A} \supset \mathrm{B}$. Let us suppose that B_1 is a set closed under B but not under A. The closure $\mathrm{A}(B_1) = A_1$ of B_1 under A then contains B_1 as a proper subset. We shall now select $\Gamma = \mathrm{M}_{C_1}$ as a maximal closure relation with the principal set $C_1 \neq 0$ taken such that

$$A_1 \cdot C_1 = B_1 .$$

This means that C_1 has the form

$$C_1 = B_1 + A_1'$$

where A_1' may be an arbitrary set contained in $S - A_1$. According to theorem 14 one must have

$$B_1 = A_1 \cdot C_1 = A_1 \cdot (\mathrm{B} \cup \Gamma)(B_1).$$

But since one only has the possibility

$$(\mathrm{B} \cup \Gamma)(B_1) = C_1$$

it follows that C_1 is closed under B. Thus we have that for every set B_1 closed under B but not under A all sets

$$C_1 = B_1 + A_1', \qquad A_1' \subset S - \mathrm{A}(B_1)$$

must be closed under B.

When this condition is satisfied the Dedekind law will hold for an arbitrary

closure relation Γ. If namely for some set C_1 which is closed under Γ and some set A_1 closed under Α the intersection

$$B_1 = A_1 \cdot C_1$$

is closed under Β then C_1 has the form

$$C_1 = B_1 + A_1'$$

where A_1' is contained in $S - \mathrm{A}(B_1)$ and C_1 is closed also under Β so that the conditions of theorem 14 are immediately satisfied.

THEOREM 18. *The necessary and sufficient condition that the Dedekind relation* (3) *hold for a fixed pair of closure relations* $\mathrm{A} \supset \mathrm{B}$ *and an arbitrary* Γ *is that for every set* B_1 *which is closed under* Β *but not under* Α *all sets of the form*

$$C_1 = B_1 + A_1', \qquad A_1' \subset S - \mathrm{A}(B_1)$$

be closed under B.

There are a great number of laws related to the Dedekind law which can be analysed in the same manner and similar results can be derived without difficulty. These investigations have also been carried through for the distributive law

$$\mathrm{A} \cap (\mathrm{B} \cup \Gamma) = (\mathrm{A} \cap \mathrm{B}) \cup (\mathrm{A} \cap \Gamma)$$

and its dual but the results shall not be stated.

In connection with the Dedekind law one can also investigate when the analogue of the algebraic law of isomorphism holds. Let A and B denote two elements in some structure. The law of isomorphism states that there shall exist a structure isomorphism between the two quotient structures

$$\mathrm{A} \cup \mathrm{B}/\mathrm{A} \cong \mathrm{B}/\mathrm{A} \cap \mathrm{B}. \tag{8}$$

In algebraic systems in which this structure isomorphism holds it is usually established by means of the so-called regular *structure correspondence*, which may be defined as follows. Let us denote by X and Y arbitrary elements in the two quotient structures so that

$$\mathrm{A} \cup \mathrm{B} \supset \mathrm{X} \supset \mathrm{A}, \qquad \mathrm{B} \supset Y \supset \mathrm{A} \cap \mathrm{B}. \tag{9}$$

The regular structure correspondence is then defined by

$$\mathrm{X} \rightarrow \mathrm{B} \cap \mathrm{X}, \qquad \mathrm{Y} \rightarrow \mathrm{A} \cup Y. \tag{10}$$

It is not difficult to prove that these correspondences (10) establish a structure isomorphism (8) if and only if for every X and Y satisfying (8) one has

$$\mathrm{X} = \mathrm{A} \cup (\mathrm{B} \cap \mathrm{X}), \qquad Y = \mathrm{B} \cap (\mathrm{A} \cup Y). \tag{11}$$

In the case of the structure of closure relations the two conditions (11) can be analysed by means of the results obtained in the theorems 15 and 16 and one obtains directly:

THEOREM 19. *The necessary and sufficient condition for the regular structure correspondence* (10) *to define a structure isomorphism* (8) *is that any intersection* $A_1 \cdot B_1$ *of sets closed under* A *and* B *be closed either under* A *or under* B *and also be representable in the form*

$$A_1 \cdot B_1 = A_1 \cdot (\mathrm{A} \cup \mathrm{B})(B_1).$$

4. The product of closure relations

The result of applying one closure relation after the other shall be called the *product* of the two closure relations. If Γ_1 and Γ_2 are the two given closure relations we shall write

$$\Gamma_1 \times \Gamma_2(A) = \Gamma_1(\Gamma_2(A))$$

The correspondence

$$A \rightarrow \Gamma_1(\Gamma_2(A))$$

is an increasing and order preserving correspondence, but it is not always idempotent, so that the product of two closure relations is usually not a closure relation.

Under special conditions the product may be a closure relation and we mention first the following simple case:

THEOREM 20. *When* $\Gamma_1 \supset \Gamma_2$ *are two closure relations, one containing the other, one has*

$$\Gamma_1 \times \Gamma_2 = \Gamma_2 \times \Gamma_1 = \Gamma_1$$

PROOF: Since $\Gamma_1 \supset \Gamma_2$ one has

$$\Gamma_2(\Gamma_1(A)) = \Gamma_1(A)$$

because $\Gamma_1(A)$ is also closed under Γ_2. To prove the other half of the theorem we observe that from

$$\Gamma_1(A) \supset \Gamma_2(A)$$

follows

$$\Gamma_1(\Gamma_1(A)) = \Gamma_1(A) \supset \Gamma_1(\Gamma_2(A)) \supset \Gamma_1(A)$$

so that

$$\Gamma_1(A) = \Gamma_1(\Gamma_2(A)).$$

Among the further properties of the product of two closure relations we mention the inclusions

$$\Gamma_1 \cup \Gamma_2(A) \supset \Gamma_1 \times \Gamma_2(A) \supset \Gamma_1(A) + \Gamma_2(A) \supset \Gamma_1 \cap \Gamma_2(A). \tag{12}$$

Only the first of these needs any proof and this follows from

$$\Gamma_1 \cup \Gamma_2(A) \supset \Gamma_2(A)$$

by taking the Γ_1-closure of both sides and applying theorem 20.

Let us say that the closure relation Γ contains the product $\Gamma_1 \times \Gamma_2$ if

$$\Gamma(A) \supseteqq \Gamma_1 \times \Gamma_2(A)$$

for every subset A. In this case one sees that

$$\Gamma(A) \supseteqq \Gamma_1(A), \qquad \Gamma(A) \supseteqq \Gamma_2(A)$$

so that we have shown:

THEOREM 21. *The union $\Gamma_1 \cup \Gamma_2$ is the smallest closure relation containing the product $\Gamma_1 \times \Gamma_2$.*

When this result is combined with (12) one obtains:

THEOREM 22. *The necessary and sufficient condition that the product $\Gamma_1 \times \Gamma_2$ of two closure relations be a closure relation is that*

$$\Gamma_1 \times \Gamma_2 = \Gamma_1 \cup \Gamma_2 .$$

This in turn leads to another criterion:

THEOREM 23. *The necessary and sufficient condition for $\Gamma_1 \times \Gamma_2$ to be a closure relation is that the Γ_1-closure of any Γ_2-closed set be Γ_2-closed.*

PROOF: If

$$\Gamma_1(\Gamma_2(A)) = \Gamma_2(B)$$

one has

$$\Gamma_2\Gamma_1\Gamma_2(A) = \Gamma_2(B) = \Gamma_1\Gamma_2(A)$$

hence

$$\Gamma_1\Gamma_2\Gamma_1\Gamma_2(A) = \Gamma_1\Gamma_2(A)$$

so that the product $\Gamma_1 \times \Gamma_2$ is an idempotent relation and therefore according to a previous remark, a closure relation. Conversely let $\Gamma_1 \times \Gamma_2$ be a closure relation, hence according to theorem 22

$$\Gamma_1 \times \Gamma_2 = \Gamma_1 \cup \Gamma_2$$

and the condition of theorem 23 is satisfied.

If both products $\Gamma_1 \times \Gamma_2$ and $\Gamma_2 \times \Gamma_1$ shall be closure relations one obtains according to theorem 22

$$\Gamma_1 \times \Gamma_2 = \Gamma_2 \times \Gamma_1 = \Gamma_1 \cup \Gamma_2$$

hence the order of the factors in the product is immaterial and we say that the two closure relations *commute*. Conversely if Γ_1 and Γ_2 do commute one has

$$(\Gamma_1 \times \Gamma_2) \times (\Gamma_1 \times \Gamma_2) = (\Gamma_1 \times \Gamma_1) \times (\Gamma_2 \times \Gamma_2) = \Gamma_1 \times \Gamma_2$$

so that $\Gamma_1 \times \Gamma_2$ is an idempotent relation, hence a closure relation. Thus we can say:

The necessary and sufficient condition for both products $\Gamma_1 \times \Gamma_2$ and $\Gamma_2 \times \Gamma_1$ of two closure relations to be closure relations is that Γ_1 and Γ_2 commute.

When Γ_1 and Γ_2 commute the Γ_1-closure of any Γ_2-closed set is Γ_2-closed and similarly the Γ_2-closure of any Γ_1-closed set is Γ_1-closed.

5. Automorphisms and characterization of the structure of closure relations

Let us consider briefly the problem of determining certain characteristic properties of the structure of closure relations so that any structure with these properties is isomorphic to a structure of all closure relations over some set.

We have already observed that every closure relation is the intersection of the maximal closure relations in which it is contained. This leads us to investigate particularly the interrelation between various maximal closure relations. The cross-cut $M_A \cap M_B$ is a closure relation with the closed sets

$$\{0, S, A, B, A \cdot B\}$$

where some of these sets may coincide. This shows that for a pair of maximal closure relations there can be at most one other maximal closure relation namely $M_{A \cdot B}$ such that

$$M_{A \cdot B} \supset M_A \cap M_B .$$

Furthermore there are two other, usually non-maximal closure relations with the closed sets

$$\{0, S, A, A \cdot B\}, \{0, S, B, A \cdot B\}$$

containing the cross-cut of M_A and M_B. Among the maximal closure relations one can introduce a partial order by writing

$$M_A > M_B \tag{13}$$

whenever $A > B$. This partial order can also be introduced in a purely structural manner by saying that (13) holds if and only if there exists another maximal closure relation M_C such that

$$M_B \supset M_A \cap M_C .$$

Let us also mention that there exist minimal elements M_a containing no other maximal closure relation in the sense of (13). Every maximal closure relation M_A is uniquely determined by the set of all such minimal M_a for which

$$M_A > M_a .$$

Clearly the set of all maximal closure relations by the inclusion relation (13) is a system isomorphic with the set of all subsets of S. This may also be expressed structurally according to the *Tarški-Stone* theorem by saying that the set of all maximal closure relations forms a completely distributive complete Boolean algebra with respect to the inclusion (13).

From these remarks one sees that if a structure is to be isomorphic to a structure of all closure relations over a set, there must exist maximal elements and to any pair of maximal elements M_1 and M_3 there can exist at most one other maximal element M_2 such that

$$M_2 \supset M_1 \cap M_3 .$$

In this case we shall write $M_1 > M_2$. One must then impose such axiomatic conditions on the maximal elements of the structure that this defines a partial order of the set of all maximal elements, and furthermore such that this partially ordered set is a complete Boolean algebra. Thus by the theorem of *Tarski-Stone* just mentioned every maximal element becomes associated with a unique subset of some set S and conversely. Finally one must postulate that every element Γ in the structure is the cross-cut of the maximal elements in which it is contained, so that Γ becomes associated with a family of sets consisting of all the subsets of S which belong to the various maximal elements containing Γ. We shall not go into further details of such a theory. It may only be mentioned that a somewhat similar characterisation has been carried out by the author for the case of the structure of all equivalence relations over a set (Ore [1]).

To conclude let us mention one result which is almost an immediate consequence of the preceding remarks:

THEOREM 24. *The group of automorphisms of the structure Σ of all closure relations defined over a set S consists of all one-to-one correspondences of the set S to itself.*

PROOF: Clearly any one-to-one correspondence of the set S represents an automorphism of Σ. Conversely let us consider some automorphism α of Σ. Under α any maximal closure relation M_A must be transformed into another maximal closure relation $M_{A'}$. Those maximal closure relations M_a in which the principal set $A = a$ is a single element have also been characterized structurally so that α must transform each M_a into some other closure relation which we can denote by $M_{a\beta}$. This defines β as a one-to-one correspondence of the set S and the automorphism $\beta^{-1}\cdot\alpha$ is seen to leave all M_a fixed. Since every M_A is uniquely determined by the set of M_a such that $M_A > M_a$ it follows that every maximal closure relation, hence every closure relation remains fixed under the automorphism $\beta^{-1}\cdot\alpha$. Thus we see that α coincides with the transformation β of the set S.

6. Homomorphisms of the structure of closure relations

Let us turn to the determination of the homomorphisms of the structure of closure relations. We shall recall that a *homomorphism* α is a correspondence between the structure Σ and another structure Σ^α such that every element in Σ^α is the image of at least one element in Σ and such that

$$\Gamma_1 \to \Gamma_1^\alpha, \qquad \Gamma_2 \to \Gamma_2^\alpha$$

shall imply

$$(14) \qquad \begin{aligned} (\Gamma_1 \cap \Gamma_2)^\alpha &= \Gamma_1^\alpha \cap \Gamma_2^\alpha \\ (\Gamma_1 \cup \Gamma_2)^\alpha &= \Gamma_1^\alpha \cup \Gamma_2^\alpha . \end{aligned}$$

A homomorphism α is said to be *complete* when the relations (14) hold for an arbitrary finite or infinite number of components.

Let us consider first the so-called *modular homomorphisms*. We denote by

Δ some fixed closure relation and define a correspondence α of Σ to the substructure of all closure relations contained in Δ by putting

$$(15) \qquad \Gamma \rightarrow \Gamma^{\alpha} = \Gamma \cap \Delta.$$

Clearly this correspondence satisfies the first condition in (14). In order that it shall satisfy the second condition it is necessary and sufficient that for every pair of closure relations Γ_1 and Γ_2 one shall have

$$(16) \qquad (\Gamma_1 \cap \Delta) \cup (\Gamma_2 \cap \Delta) = (\Gamma_1 \cup \Gamma_2) \cap \Delta.$$

When this condition (16) is fulfilled for all Γ_1 and Γ_2 we shall say that Δ is a *distributive closure relation* and the resulting homomorphism defined by (15) is a *modular homomorphism*.

All the distributive closure relations may be determined as follows: We assume that the two trivial cases $\Delta = V$ and $\Delta = I$ are excluded and denote by D some set different from O and S which is closed under Δ. The two closure relations

$$\Gamma_1 = \mathrm{M}_{C_1}, \qquad \Gamma_2 = \mathrm{M}_{C_2}$$

are taken as maximal closure relations whose principal sets C_1 and C_2 are selected in the following manner: We take C_1 as an arbitrary subset of D and define

$$C_2 = C_1 + K_2$$

where K_2 is some non-void subset of $S-D$. In this case the set

$$C_1 = D \cdot C_1 = D \cdot C_2$$

is closed both under $\Gamma_1 \cap \Delta$ and $\Gamma_2 \cap \Delta$ hence under the left-hand side of (16). On the other hand the union $\Gamma_1 \cup \Gamma_2$ is the universal closure relation V so that the right-hand side in (16) is Δ. Thus we see that C_1 must be closed under Δ. We have shown therefore that a distributive closure relation must have the property that every subset of a closed set $D \neq S$ is closed.

Conversely it is not difficult to verify that this condition is sufficient for a closure relation to be distributive. Let Γ_1 and Γ_2 be two arbitrary closure relations and Δ a closure relation with the property that any subset of a closed set $D \neq S$ is closed. In this case one sees that the sets closed under $\Gamma_1 \cap \Delta$ are those which are closed under Γ_1 or Δ and similarly the sets closed under $\Gamma_2 \cap \Delta$ are closed under Γ_2 or Δ. Consequently the sets closed simultaneously under $\Gamma_1 \cap \Delta$ and $\Gamma_2 \cap \Delta$ are either sets closed under Δ or sets closed both under Γ_1 and Γ_2 so that the relation (16) holds. The relation analogous to (16) for an arbitrary finite or infinite number of closure relations Γ_i

$$(17) \qquad \bigvee_i (\Gamma_i \cap \Delta) = (\bigvee_i \Gamma_i) \cap \Delta$$

is also seen to hold when Δ is distributive. A closure relation Δ for which (17) always holds may be called *completely distributive*. Such completely distributive

closure relations are seen to define complete homomorphisms of the structure. We have shown therefore:

THEOREM 25. *The necessary and sufficient condition for a closure relation* Δ *to be distributive is that every subset of a closed set* $D \neq S$ *be closed. In this case* Δ *is completely distributive and the corresponding modular homomorphism is complete.*

The homomorphism (15) associates with each closure relation Γ the closure relation $\Gamma \cap \Delta$ obtained from Γ by adjoining to the closed sets $\Gamma(A)$ those subsets of $\Gamma(A)$ which are closed under Δ.

In connection with the distributive relation (16) we shall also analyze the dual relation

$$(\Gamma_1 \cup \Delta) \cap (\Gamma_2 \cup \Delta) = (\Gamma_1 \cap \Gamma_2) \cup \Delta. \tag{18}$$

Any closure relation Δ which satisfies this relation (18) for every pair of closure relations Γ_1 and Γ_2 may be called *dually distributive*. Any dually distributive closure relation defines a *dual modular homomorphism* through the correspondence

$$\Gamma \to \Gamma^{\alpha} = \Gamma \cup \Delta. \tag{19}$$

Let us determine all dually distributive closure relations Δ. Again we omit the two trivial cases $\Delta = V$ and $\Delta = I$. We denote by D some set different from S and O which is closed under Δ. Furthermore let C_1 and C_2 be two different sets such that

$$D = C_1 \cdot C_2, \qquad C_1 \neq D, \qquad C_2 \neq D. \tag{20}$$

Two such sets C_1 and C_2 can always be found provided D is not a maximal subset $S\text{-}d'$ of S. One of them can be taken as an arbitrary set containing D. Next we define Γ_1 and Γ_2 as maximal closure relations

$$\Gamma_1 = M_{c_1}, \qquad \Gamma_2 = M_{c_2}$$

with the principal sets C_1 and C_2. According to (20) the set D is closed under $\Gamma_1 \cap \Gamma_2$ and Δ, hence under their union $(\Gamma_1 \cap \Gamma_2) \cup \Delta$. Consequently D must also be closed under the left-hand side of (18). But aside from O and S the closure relations $\Gamma_1 \cup \Delta$ and $\Gamma_2 \cup \Delta$ can only have the closed sets C_1 and C_2 and these only when they are also closed under Δ. Thus in (20) both C_1 and C_2 are closed under Δ and we have as a necessary condition for a closure relation Δ to be dually distributive that every set C containing a closed set $D \neq O$ of Δ must itself be closed under Δ.

Conversely when this condition is satisfied the relation (18) will hold for all pairs of closure relations Γ_1 and Γ_2. If namely C_1 and C_2 are sets closed under Γ_1 and Γ_2 respectively such that their intersection $D = C_1 \cdot C_2$ is closed under Δ then C_1 and C_2 must also be closed under Δ so that D is closed under the left-hand side of (18). Also in this case one finds

$$\bigwedge_i (\Gamma_i \cup \Delta) = (\bigwedge_i \Gamma_i) \cup \Delta$$

for an arbitrary set of closure relations Γ_i so that a dually distributive closure relation is at the same time dually completely distributive and the homomorphism (19) is complete.

The condition for a closure relation Δ to be dually distributive can be expressed in a somewhat different manner. When D is a set closed under Δ every maximal set $S-d'$ containing D is closed. Consequently the sets closed under Δ may be generated by the sets $S-d'$ where d' runs through all isolated points d' of Δ. Thus we see that the sets closed under Δ are those which contain the fixed set

$$O'_\Delta = S - O_\Delta$$

where O_Δ is the set of isolated points under Δ. We state therefore:

THEOREM 26. *The necessary and sufficient condition that a closure relation be dually distributive is that its closed sets consist of all sets containing some fixed arbitrary set O'_Δ. Such closure relations are always dually completely distributive.*

We shall proceed to the determination of all complete homomorphisms of the structure of closure relations. The solution of this problem is derived as a consequence of an analysis of the homomorphisms of a general class of structures. Let us assume first that Σ is some complete structure and α a complete homomorphism of Σ. All those elements a in Σ which have the same image a^α under α are seen to form a complete substructure Σ_{a^α} of Σ. We denote by a_1 and a_2 respectively the universal element and the zero element of Σ_{a^α} so that all elements a having the same image a^α lie between a_1 and a_2

$$a_1 \supset a \supset a_2 .$$

In particular let e_1 be the greatest element in Σ having the same image e^α as the zero element e of Σ. Since a and $a \cup e_1$ will always have the same image one sees that for any a

$$a_1 \supset e_1 . \tag{21}$$

Next we suppose that Σ is a point structure so that every element a is the union of the set A_q of those points p_a which are contained in it

$$a = \bigvee_{p_a \subset A_a} p_a$$

From $a_1 \supset a_2$ one concludes $A_{a_1} \supset A_{a_2}$ so that one can write

$$A_{a_1} = A_{a_2} + C$$

where C is the set of points p_c contained in a_1 but not in a_2. We denote by

$$c = \bigvee_{p_c \supset C} p_c$$

the union of all points p_c contained in C, so that one has

$$a_1 = a_2 \cup c. \tag{22}$$

For any point p_c in c obviously

$$a_2 \cap p_c = e, \qquad a_2 \subset a_2 \cup p_c \subset a_1 .$$

The homomorphism α can now be applied to these relations. One obtains

$$a^\alpha \cap p_c^\alpha = e^\alpha, \qquad a^\alpha = a^\alpha \cup p_c^\alpha$$

and from these two relations one concludes

$$p_c^\alpha = e^\alpha$$

so that all points p_c in c have the same image e^α as e. But since α is a complete homomorphism and since c is the union of all p_c one obtains further

$$c^\alpha = e^\alpha.$$

This shows that $e_1 \supset c$ where e_1 is the maximal element with the same image as e. Thus one concludes finally from (21) and (22) that

$$a_1 = a_2 \cup e_1$$

so that two elements a' and a'' have the same image a^α if and only if

$$a' \cup e_1 = a'' \cup e_1$$

and we have proved our main result:

THEOREM 27. *In a complete point structure all complete homomorphisms are dually modular.*

A structure in which the finite chain condition is satisfied is always complete and it is easily seen that its homomorphisms must also be complete so that we can state:

THEOREM 28. *In a point structure satisfying the finite chain condition all homomorphisms are dually modular.*

It had already been established that the structure of all closure relations over a set is a dual point structure, i.e. every element is the cross-cut of the maximal closure relations in which it is contained. From the dual of theorem 27 one obtains therefore:

THEOREM 29. *In the structure of all closure relations over a set all complete homomorphisms are modular.*

I have not been able to determine all non-complete homomorphisms although certain types of such homomorphisms have been found, for instance by the following construction: Let Σ_0 be some substructure of Σ. We shall introduce an equivalence relation in Σ by writing

$$\Gamma_1 \equiv \Gamma_2 \pmod{\Sigma_0}$$

provided there exists an element Δ_0 in Σ_0 such that (Ore [1])

$$\Gamma_1 \cap \Delta_0 = \Gamma_2 \cap \Delta_0 .$$

One finds without difficulty that this equivalence defines a homomorphism of Σ if and only if to every pair Γ_1 and Γ_2 of closure relations and every Δ_0 in Σ_0 one can find another closure relation Δ_0' in Σ_0 such that

$$((\Gamma_1 \cap \Delta_0) \cup (\Gamma_2 \cap \Delta_0)) \cap \Delta_0' = (\Gamma_1 \cup \Gamma_2) \cap \Delta_0'. \tag{23}$$

This condition is fulfilled for a substructure of Σ_0 which has the property that

to every Δ_0 in Σ_0 also every Δ_0^* is in Σ_0, where Δ^* is the distributive closure relation obtained from Δ_0 through the adjunction of all sets D_0' contained in the closed sets $D_0 \neq S$ of Δ_0. Whether all substructures Σ_0 satisfying (23) can be obtained this way is not known. Nor is it known whether all non-complete homomorphisms are of this generalized modular form.

7. Point closures

In closure relations some assumption is usually made about the closure of a point. The simplest requirement is of course:

Closed points. Every point is closed.

The general closure relations in which every point is closed are also seen to form a complete dual point structure Σ'. This structure Σ' is homomorphic to the structure Σ of all general closure relations and it consists of all closure relations contained in the closure relation Γ_0 in which the points are the only closed sets besides O and S. In this structure Σ' the maximal closure relations M_A' are those whose closed sets are O, S the set A and all points. The structural theory of these closure relations is practically the same as for general closure relations.

Instead of assuming that every point is a closed set one can make the weaker assumption that every point is determined by its closure:

Point determination. When $\bar{a} = \bar{b}$ then $a = b$.

This condition can also be stated in a different form which is often more convenient in its applications. To arrive at this reformulation let us recall first that a complete field of sets over S is a complete ring $\mathfrak{R}$ with the additional property that the complement $S-A$ of every set A in $\mathfrak{R}$ is also in the ring. Let us also recall that there exists a one-to-one correspondence between the complete field of sets over S and the partitions $\mathfrak{P}(B_i)$ of S such that each complete field $\mathfrak{R}$ consists of the sums of the blocks B_i of a particular partition $\mathfrak{P}$ of S. (Ore [1])

Let $\mathfrak{F}(A_m)$ be a family of sets covering the set S, i.e. every element a is contained in at least one set A_m. We assume that the indices m of the sets A_m in the family run through an index set M whose cardinal number may be arbitrary. Since every such family of sets $\mathfrak{F}(A_m)$ generates a least field of sets over S in which it is contained, there is also associated a unique partition $\mathfrak{P}(B_{M_1})$ of S with every family $\mathfrak{F}(A_m)$. The blocks B_{M_1} in this partition may be obtained by the following construction. Let a be an element in S and M_1 the set of all indices m_1 for which the set A_{m_1} contains a. Similarly $M_2 = S-M_1$ is the set of those indices m_2 for which A_{m_2} does not contain a. Consequently the first of the two sets

$$\prod_{m_1 \subset M_1} A_{m_1}, \qquad \sum_{m_2 \subset M_2} A_{m_2}$$

contains a while the second does not. Thus

$$B_{M_1} = \prod_{m_1 \subset M_1} A_{m_1} - \prod_{m_1 \subset M_1} A_{m_1} \cdot \sum_{m_2 \subset S-M_1} A_{m_2} \tag{24}$$

is a set containing a. Two sets B_{M_1} determined by two different elements a_1 and a_2 are seen to be either disjoint or identical so that when M_1 runs through

all subsets of M one obtains a partition $\mathfrak{P}(B_{M_1})$ of S. The blocks (24) obviously belong to the field generated by $\mathfrak{F}(A_m)$. On the other hand from

$$A_{m_1} \cdot B_{M_1} = B_{M_1}, \qquad A_{m_2} \cdot B_{M_1} = 0$$

follows that every set A_m is the sum of such blocks so that the partition $\mathfrak{P}(B_{M_1})$ is the partition associated with the field generated by $\mathfrak{F}(A_m)$.

After these preparatory remarks we can state:

THEOREM 30. *The necessary and sufficient condition that in a closure relation every point be uniquely determined by its closure is that the family of closed sets define a complete partition of S in which every block consists of a single point.*

PROOF: If a and b are two different points with the same closure then they are both contained in the same closed sets A_{m_1} so that a and b must belong to the same block B_{M_1} in (24). On the other hand if a and b have different closures then one cannot have simultaneously

$$\bar{a} \supset b, \qquad \bar{b} \supset a$$

because the closed set $\bar{a} \cdot \bar{b}$ would contain both a and b against the fact that $\bar{a}$ and $\bar{b}$ are the smallest closed sets containing a and b respectively. Thus when a and b have different closures they belong to different blocks B_{M_1} in (24). Theorem 30 follows as an immediate consequence.

From theorem 30 one concludes that when Γ_1 and Γ_2 are closure relations whose points are determined by their closures, the union $\Gamma_1 \cup \Gamma_2$ whose closed sets are those common to Γ_1 and Γ_2 will usually not have this property. This shows that for the set of all closure relations in which every point is determined by its closure the natural definition of a union breaks down.

YALE UNIVERSITY.

BIBLIOGRAPHY

GARRETT BIRKHOFF [1]: *On the combination of topologies.* Fundamenta Math. v. 29 (1936) pp. 156–166.

E. H. MOORE: *Introduction to a form of general analysis.* New Haven Mathematical Colloquium 1910.

OYSTEIN ORE [1]: *Theory of equivalence relations.* Duke Math. Journ. v. 9 (1942) pp. 573–627.

ANNALS OF MATHEMATICS
Vol. 44, No. 3, July, 1943

SOME REMARKS ON ALGEBRAS OVER AN ALGEBRAICALLY CLOSED FIELD*

BY C. NESBITT AND W. M. SCOTT

(Received October 11, 1942)

The theory of rings with radicals is an interesting and far reaching problem of modern algebra.[1] In this paper we have examined some aspects of algebras which may have radicals and whose coefficient fields are algebraically closed. Some of the methods employed clearly could be used for less restricted algebras, but a full extension of the results requires the solution of a number of problems still under investigation.[2] The authors feel that the theory of algebras over an algebraically closed field has some interest and value in itself, particularly in view of its immediate application to the representation theory of finite groups. Moreover, this restricted case is a valuable testing ground for theorems for more general rings.

In the first part of the paper we have studied the concept of basic algebra. The basic algebras are semi-primitive subalgebras (i.e. modulo their radicals are direct sums of division algebras, in our case, direct sums of fields) which for the algebras under discussion play a role in some respects analogous to that of division algebras for simple algebras. Related to the basic algebras are the Cartan basis systems,[3] and systems of elementary modules.[4] The commutator algebras of matrix representations of an algebra, or what is equivalent, the algebras of homomorphisms of the related representation spaces, can be analyzed in a rather simple manner.

We shall say that a linear function φ of an algebra $\mathfrak{a}$ is symmetric, if for every $\alpha, \beta \in \mathfrak{a}$, $\varphi(\alpha\beta) = \varphi(\beta\alpha)$. In the case where $\mathfrak{a}$ is over an algebraically closed field, and is also semisimple, the characters of the irreducible representations of $\mathfrak{a}$ form a complete set of symmetric functions of $\mathfrak{a}$. When $\mathfrak{a}$ has a radical, this is no longer true. In Part 2 of the paper, we discuss symmetric functions of algebras with radical.

In Part 3 of the paper the regular representations are written in terms of elementary modules.

---

* The following includes the second part of a dissertation written by W. M. Scott under the direction of C. Nesbitt, and accepted by the University of Michigan in January, 1941. [12], in the bibliography at the end, is the first part of the dissertation. A remark by R. M. Thrall, who read the dissertation, lead to the development of Part 1 of this paper.

[1] For some of the more recent developments, see the bibliography.

[2] B. Vinograde has done some work in this direction. See Bulletin of the American Mathematical Society, vol. 48, No. 5, Abstract 167.

[3] For a discussion of the Cartan basis system, see [11], §2.

[4] Elementary modules are defined in [12], §3.

I. Basic Algebras†

1. Main concepts

Let $\mathfrak{a}$ be an algebra with unit element over an algebraically closed field K, and let

$$\mathfrak{a} = \mathfrak{a}^* + \mathfrak{n} \tag{1}$$

be a splitting of $\mathfrak{a}$ into a direct sum of a semisimple subalgebra $\mathfrak{a}^*$ and the radical $\mathfrak{n}$ of $\mathfrak{a}$.[5] We shall denote by

$$\mathfrak{a}^* = \mathfrak{a}_1^* + \cdots + \mathfrak{a}_k^* \tag{2}$$

the unique splitting of $\mathfrak{a}^*$ into a direct sum of simple invariant subalgebras, and the components of $\alpha \in \mathfrak{a}$ under the splittings (1), (2) by

$$\begin{aligned} \alpha &= \alpha^* + \eta \\ \alpha^* &= \alpha_1^* + \cdots + \alpha_k^* \end{aligned} \tag{3}$$

The radical $\mathfrak{n}$ possesses the series

$$\mathfrak{n} \supset \mathfrak{n}^2 \supset \cdots \supset \mathfrak{n}^{l-1} \supset \mathfrak{n}^l = 0 \tag{4}$$

of invariant subalgebras of $\mathfrak{a}$.

We shall call a vector space V which has a ring $\mathfrak{b}$ as left operator system and a ring $\mathfrak{c}$ as right operator system a $(\mathfrak{b}, \mathfrak{c})$ space if, in addition to the usual operator relations, for every $\xi \in \mathfrak{b}$, $X \in V$, $\zeta \in \mathfrak{c}$ the associativity condition

$$(\xi X)\zeta = \xi(X\zeta) \tag{5}$$

is satisfied. The condition (5) implies that each $\zeta \in \mathfrak{c}$ determines a $\mathfrak{b}$-homomorphism, $X \to X\zeta$, of V considered as a $\mathfrak{b}$ left space, and similarly each $\xi \in \mathfrak{b}$ determines a $\mathfrak{c}$-homomorphism of V considered as a $\mathfrak{c}$ right space. If we use functional notation to denote $\mathfrak{b}$-homomorphisms of V, for example, write $X \to f(X) = X\zeta$, $X \to f_1(X) = X\zeta_1$, then $f_1 f(X) = f_1(f(X)) = f_1(X\zeta) = X\zeta\zeta_1$. Thus between elements of $\mathfrak{c}$ and elements of the $\mathfrak{b}$-homomorphism ring $\mathfrak{m}$ of V we have relations of form $\zeta \to f$, $\zeta_1 \to f_1$, $\zeta\zeta_1 \to f_1 f$. This shows that $\mathfrak{c}$ is homomorphically mapped into the ring $\mathfrak{m}'$ inverse to $\mathfrak{m}$. We shall call $\mathfrak{m}'$ the inverse $\mathfrak{b}$-homomorphism ring of V.

Let now

$$\mathfrak{a} = \mathfrak{a}_1 \supset \mathfrak{a}_2 \supset \cdots \supset \mathfrak{a}_r \supset (0) \tag{6}$$

be a composition series for $\mathfrak{a}$ considered as an $(\mathfrak{a}, \mathfrak{a})$ space. The analysis of the algebra $\mathfrak{a}$ we are considering is simplified by the following theorem.

---

† The notion of basic algebra has been discussed in lectures by R. Brauer.

[5] For the proof that such a splitting exists, see, for example, [1], p. 47.

1.1 *Let V be an irreducible* $(\mathfrak{a}, \mathfrak{a})$ *space.*[6] *Then V is also an irreducible* $(\mathfrak{a}^*, \mathfrak{a}^*)$ *space.*

PROOF: $\mathfrak{n}V$ is an $(\mathfrak{a}, \mathfrak{a})$ subspace of V. Then $\mathfrak{n}V = (0)$ or V. If $\mathfrak{n}V = V$, then $V = \mathfrak{n}^\tau V$ for $\tau = 1, 2, \cdots$ and this implies $V = (0)$, in which case the theorem is trivial. The same remarks hold for $V\mathfrak{n}$. If $\mathfrak{n}V = V\mathfrak{n} = (0)$, then for $\alpha = \alpha^* + \eta \in \mathfrak{a}$, $X \in V$, $\alpha X = (\alpha^* + \eta)X = \alpha^* X$; similarly $X\alpha = X\alpha^*$. In this case the operators α and α^* produce the same effect on V, so that V irreducible as an $(\mathfrak{a}, \mathfrak{a})$ space implies V is irreducible as an $(\mathfrak{a}^*, \mathfrak{a}^*)$ space.

It is clear that a corresponding theorem holds for spaces having $\mathfrak{a}$ as one-sided operator system.

1.2 *A composition series of* $\mathfrak{a}$ *considered as an* $(\mathfrak{a}, \mathfrak{a})$ *space is also a composition series of* $\mathfrak{a}$ *considered as an* $(\mathfrak{a}^*, \mathfrak{a}^*)$ *space.*

PROOF: Each composition factor group in the series is an irreducible $(\mathfrak{a}, \mathfrak{a})$ space and by 1.1 is then an irreducible $(\mathfrak{a}^*, \mathfrak{a}^*)$ space.

1.3 *Let V be an irreducible* $(\mathfrak{a}, \mathfrak{a})$ *space* $\neq (0)$. *Then there exists a unique index pair* (κ, λ), $\kappa, \lambda = 1, 2, \cdots$, *or k such that* $\mathfrak{a}_\kappa^* V \mathfrak{a}_\lambda^* = V$.

PROOF: By 1.1, $V = \mathfrak{a}^* V \mathfrak{a}^* = \sum_{\mu,\nu=1}^{k} \mathfrak{a}_\mu^* V \mathfrak{a}_\nu^*$. Each $\mathfrak{a}_\mu^* V \mathfrak{a}_\nu^*$ is an $(\mathfrak{a}^*, \mathfrak{a}^*)$ subspace of V, and so is either (0) or V. At least one of these subspaces is different from (0), say $\mathfrak{a}_\kappa^* V \mathfrak{a}_\lambda^* \neq (0)$. If $V = \mathfrak{a}_\kappa^* V \mathfrak{a}_\lambda^*$ contains another subspace $\mathfrak{a}_\rho^* V \mathfrak{a}_\sigma^* \neq (0)$, then $\mathfrak{a}_\rho^* V \mathfrak{a}_\sigma^* = \mathfrak{a}_\kappa^* V \mathfrak{a}_\lambda^*$, and since $\mathfrak{a}_\mu^* \mathfrak{a}_\nu^* = \delta_{\mu\nu} \mathfrak{a}_\mu^*$[7] this gives a contradiction, and the theorem is proved.

We shall say that an irreducible V is of type (κ, λ) if $V = \mathfrak{a}_\kappa^* V \mathfrak{a}_\lambda^*$.

1.4 *Let V be an irreducible* $(\mathfrak{a}, \mathfrak{a})$ *space* $\neq (0)$ *and of type* (κ, λ). *Let* $e_{\mu,ab}$ $(a, b = 1, 2, \cdots, f_\mu)$ *denote a set of matrix units for the simple algebra* $\mathfrak{a}_\mu^*$. *Then there exists a vector* $X = e_{\kappa,11} X e_{\lambda,11} \in V$ *such that* $e_{\kappa,a1} X e_{\lambda,1b}$ $(a = 1, 2, \cdots, f_\kappa;$ $b = 1, 2, \cdots, f_\lambda)$ *form a K-basis of V.*

PROOF: The unit element $e_\kappa = \sum_{a=1}^{f_\kappa} e_{\kappa,aa}$ of $\mathfrak{a}_\kappa^*$ is a left identity operator, and $e_\lambda = \sum_{b=1}^{f_\lambda} e_{\lambda,bb}$ is a right identity operator for V. If $Y \neq 0 \in V$, $e_\kappa Y e_\lambda = Y$, so that for some p, q, $e_{\kappa,pp} Y e_{\lambda,qq} \neq 0$. But $e_{\kappa,pp} Y e_{\lambda,qq} = e_{\kappa,p1}(e_{\kappa,1p} Y e_{\lambda,q1}) e_{\lambda,1q}$, so $e_{\kappa,1p} Y e_{\lambda,q1} = X \neq 0$. Moreover, $e_{\kappa,11} X e_{\lambda,11} = X$. Since $\mathfrak{a}_\kappa^* = \sum_{i,j=1}^{f_\kappa} K e_{\kappa,ij}$, $\mathfrak{a}_\lambda^* = \sum_{h,k=1}^{f_\lambda} K e_{\lambda,hk}$ it follows that $V' = \sum_{a=1}^{f_\kappa} \sum_{b=1}^{f_\lambda} K e_{\kappa,a1} X e_{\lambda,1b}$ is an $(\mathfrak{a}^*, \mathfrak{a}^*)$ subspace $\neq (0)$ of V, and since V is irreducible, then $V' = V$. This shows that the elements $e_{\kappa,a1} X e_{\lambda,1b}$ $(a = 1, 2, \cdots, f_\kappa, b = 1, 2, \cdots, f_\lambda)$ give a K-basis for V.

1.5 *An irreducible* $(\mathfrak{a}, \mathfrak{a})$ *space* $\neq (0)$ *and of type* (κ, λ) *is the direct sum of* f_λ *irreducible* $\mathfrak{a}$ *left spaces and is also the direct sum of* f_κ *irreducible* $\mathfrak{a}$ *right spaces.*

PROOF: $V = V_1 + \cdots + V_{f_\lambda}$ where V_d, $(d = 1, 2, \cdots, f_\lambda)$ is the $\mathfrak{a}$-left space with basis $e_{\kappa,a1} X e_{\lambda,1d}$ $(a = 1, 2, \cdots, f_\kappa)$. V_d is evidently an irreducible $\mathfrak{a}_\kappa^* = \sum_{i,j=1}^{f_\kappa} K e_{\kappa,ij}$ space and so is an irreducible $\mathfrak{a}$ space. Similarly, $V = W_1 + \cdots + W_{f_\kappa}$, where W_c is the irreducible $\mathfrak{a}$ right space with basis $e_{\kappa,c1} X e_{\lambda,1b}$ $(b = 1, 2, \cdots, f_\lambda)$.

It will follow from a later result (cf. 4.1) that for V as in 1.5 $\mathfrak{a}_\lambda^*$ is isomorphic to the inverse $\mathfrak{a}_\kappa^*$-homomorphism ring of V.

---

[6] An irreducible $(\mathfrak{a}, \mathfrak{a})$ space is one which does not contain a proper $(\mathfrak{a}, \mathfrak{a})$ subspace.

[7] $\delta_{\mu\nu}$ (Kronecker delta) $= 0$, $\mu \neq \nu$; $\delta_{\mu\mu} = 1$.

In the following we shall denote the type of the composition factor group $\mathfrak{a}_u/\mathfrak{a}_{u+1}$ by (κ_u, λ_u), $(u = 1, 2, \cdots, r)$, and residue classes by $< \cdots >$.

1.6 *Let* $<\beta_u> \neq <0>$ *be chosen from* $\mathfrak{a}_u/\mathfrak{a}_{u+1}$ $(u = 1, 2, \cdots, r)$ *such that* $\beta_u = e_{\kappa_u,11}\beta_u e_{\lambda_u,11}$.
Then the elements

$$e_{\kappa_u,a1}\beta_u e_{\lambda_u,1b} \qquad \begin{matrix} u = 1, 2, \cdots, r \\ a = 1, 2, \cdots, f_{\kappa_u} \\ b = 1, 2, \cdots, f_{\lambda_u} \end{matrix} \tag{7}$$

form a K*-basis of* $\mathfrak{a}$.

PROOF: Since $<\beta_u> \neq <0>$, $u = 1, 2, \cdots, r$ it follows that $\beta_1, \beta_2, \cdots, \beta_r$ are linearly independent. Since by 1.4 the elements $e_{\kappa_u,a1}\beta_u e_{\lambda_u,1b}$ form a basis for $\mathfrak{a}_u/\mathfrak{a}_{u+1}$, the elements (7) give a basis for $\mathfrak{a}$.

This basis has been referred as the Cartan basis system in [5], [11], [12]. In regard to this basis an element $\alpha \in \mathfrak{a}$ may be expressed as

$$\alpha = \sum\nolimits_{u,a,b} h^u_{ab}(\alpha) e_{\kappa_u,a1}\beta_u e_{\lambda_u,1b}. \tag{8}$$

The additive group formed by the matrices $\mathfrak{H}_u(\alpha) = (h^u_{ab}(\alpha))_{ab}$ was called an *elementary module* of $\mathfrak{a}$ in [12], and denoted by $\mathfrak{H}_u$. β_u and $\mathfrak{H}_u$ will also be said to be of type (κ_u, λ_u), that is, of the same type as the corresponding composition factor group $\mathfrak{a}_u/\mathfrak{a}_{u+1}$. The number of composition factor groups in (6) which are of type (κ, λ) is denoted by $c_{\kappa\lambda}$. These $c_{\kappa\lambda}$ were called the Cartan invariants in [5], [11].

Elements β_u chosen as in 1.6 we shall call *primitive elements* of $\mathfrak{a}$. We observe that the system $\beta_1, \cdots, \beta_r$ is chosen with regard to a particular splitting (1) and a particular composition series (6).

1.7 *Let* $\beta_1, \beta_2, \cdots, \beta_r$ *be a system of primitive elements of* $\mathfrak{a}$. *Then* $\bar{\mathfrak{a}} = K\beta_1 + K\beta_2 + \cdots + K\beta_r$ *is an algebra over* K. *Any other system* $\gamma_1, \cdots, \gamma_r$ *of primitive elements of* $\mathfrak{a}$ *yields an algebra which is isomorphic to* $\bar{\mathfrak{a}}$.

PROOF: $\bar{\mathfrak{a}}$ is evidently closed under addition; it remains to be shown that it is closed under multiplication.

Let $\beta_{i_1}, \cdots, \beta_{i_m}$ denote the $m = c_{\rho\sigma}$ elements β_u of type (ρ, σ). Then every element α of $\mathfrak{a}$ such that $e_{\rho,11}\alpha e_{\sigma,11} = \alpha$ is contained in $K\beta_{i_1} + \cdots + K\beta_{i_m}$, as one sees from (8). Now let β_u, β_v be such that $\kappa_u = \rho$, $\lambda_v = \sigma$. Then $\beta_u\beta_v = e_{\rho,11}\beta_u\beta_v e_{\sigma,11}$, and so is contained in $K\beta_{i_1} + \cdots + K\beta_{i_m}$, and hence in $\bar{\mathfrak{a}}$. This shows that $\bar{\mathfrak{a}}$ is also closed under multiplication and so is a subalgebra of $\mathfrak{a}$.

Let $\gamma_1, \cdots, \gamma_r$ be a second system of primitive elements chosen with respect to the composition series (6); then $\gamma_1, \cdots, \gamma_r$ are linearly independent. Further, if γ_v is of type (ρ, σ), then $\gamma_v = e_{\rho,11}\gamma_v e_{\sigma,11}$, and by the above paragraph is contained in $K\beta_{i_1} + \cdots + K\beta_{i_m}$. It follows that $\bar{\mathfrak{a}} = K\beta_1 + \cdots + K\beta_r = K\gamma_1 + \cdots + K\gamma_r$. This shows that any two systems of primitive elements chosen with respect to the same composition series (6) and same splitting (1) yield the same algebra $\bar{\mathfrak{a}}$.

If now we have another composition series

$$(6') \qquad \mathfrak{a} = \mathfrak{b}_1 \supset \mathfrak{b}_2 \supset \cdots \supset \mathfrak{b}_r \supset (0)$$

different from (6), a theorem of Brauer's shows that we may select complete residue systems P_u for the $\mathfrak{a}_u/\mathfrak{a}_{u+1}$ and Q_v for the $\mathfrak{b}_v/\mathfrak{b}_{v+1}$ such that each P_u is the same as some Q_v.[8] Let $\delta \in P_u = Q_v$, then for some p, q, $e_{\kappa_u,1p}\delta e_{\lambda_u,q1} \neq 0$, and we may take $\beta_u = \gamma_v = e_{\kappa_u,1p}\delta e_{\lambda_u,q1}$. Then the system $\beta_1, \cdots, \beta_r$ of primitive elements with regard to (6) and the splitting (1) is the same as the system of primitive elements $\gamma_1, \cdots, \gamma_r$ with regard to (6′) and (1), and each is a basis for the algebra $\bar{\mathfrak{a}}$.

Finally, if we had another splitting

$$(9) \qquad \mathfrak{a} = \mathfrak{a}^{*\prime} + \mathfrak{n}$$

different from (1), the matrix units $e_{\kappa,ij}$ of $\mathfrak{a}^*$ are congruent modulo $\mathfrak{n}$ to matrix units $e'_{\kappa,ij}$ of $\mathfrak{a}^{*\prime}$. Then if $\beta_1, \cdots, \beta_r$ is a system of primitive elements with regard to the splitting (1), $\beta'_u = e'_{\kappa_u,11}\beta_u e'_{\lambda_u,11}$ ($u = 1, 2, \cdots, r$) form a system of primitive elements with regard to the splitting (9) and the correspondences $c_1\beta_1 + \cdots + c_r\beta_r \leftrightarrow c_1\beta'_1 + \cdots + c_r\beta'_r$, $c_u \in K$, show that the basic algebras $K\beta_1 + \cdots + K\beta_r$ and $K\beta'_1 + \cdots + K\beta'_r$ are isomorphic.

We shall call the algebra $\bar{\mathfrak{a}}$ determined by any system of primitive elements of $\mathfrak{a}$, the basic algebra of $\mathfrak{a}$. 1.6 shows that the basic algebra is unique up to isomorphism.

1.8 $\bar{\mathfrak{a}}$ *has a unit element.*

Proof: As the beginning of a composition series for $\mathfrak{a}$, we may take $\mathfrak{a} = \mathfrak{c}_1 \supset \mathfrak{c}_2 \supset \cdots \supset \mathfrak{c}_k$ where $\mathfrak{c}_\mu = \mathfrak{a}^*_\mu + \cdots + \mathfrak{a}^*_k + \mathfrak{n}$. Then $\mathfrak{c}_\mu/\mathfrak{c}_{\mu+1} \cong \mathfrak{a}^*_\mu$, and we may choose $e_{\mu,11}$ as the primitive element corresponding to this factor group. Then $\bar{e} = \sum_{\mu=1}^{r} e_{\mu,11}$ is the unit element of $\bar{\mathfrak{a}}$.

2. Connections between representations of $\mathfrak{a}$ and $\bar{\mathfrak{a}}$

In the following we shall always assume that our representations are such that the unit element of the algebra represented is an identity operator on the corresponding representation spaces. This excludes the possibility of representations with 0-constituents. Also, as it is customary and convenient, we shall consider equivalent representations to be identical.

2.1 *There is a* (1-1) *correspondence between the representations of* $\mathfrak{a}$ *and those of* $\bar{\mathfrak{a}}$. *If* $\mathfrak{A}$ *is a representation of* $\mathfrak{a}$, *and* $\bar{\mathfrak{A}}$ *is the corresponding representation of* $\bar{\mathfrak{a}}$, *then* $\mathfrak{A}$ *and* $\bar{\mathfrak{A}}$ *have corresponding structures.*

Proof: Let $\mathfrak{A}$ be a representation of $\mathfrak{a}$, and let V be the representation space of $\mathfrak{A}$. Let

$$(10) \qquad V = V_1 \supset V_2 \supset \cdots \supset V_t \supset (0)$$

be a composition series for V considered as an $\mathfrak{a}$-left space, and

[8] [3], Th. (1.2A).

let $<Y_h> \neq <0>$ be an element of V_h/V_{h+1}. Then for some ν_h and a, $e_{\nu_h,aa}Y_h \neq 0$, hence since $e_{\nu_h,aa} = e_{\nu_h,a1}e_{\nu_h,1a}$, $X_h = e_{\nu_h,1a}Y_h \neq 0$. It follows that

$$(11) \qquad <X_h> = <e_{\nu_h,11}X_h>, \quad <e_{\nu_h,21}X_h>, \cdots, <e_{\nu_h,f_{\nu_h}1}X_h>$$

is a K-basis of V_h/V_{h+1}. Repeating the process for each of the factor groups V_h/V_{h+1} we obtain a K-basis of V consisting of vectors

$$(12) \qquad \begin{aligned} X_{h,a} = e_{\nu_h,a1}X_h, \qquad & h = 1, 2, \cdots, t \\ & a = 1, 2, \cdots, f_{\nu_h} \quad \text{where} \quad X_h = e_{\nu_h,11}X_h = X_{h,1}. \end{aligned}$$

Now let $\bar{\mathfrak{a}} = K\beta_1 + \cdots + K\beta_r$, and recall that for each primitive element β_u there is an index pair (κ, λ) such that $\beta_u = e_{\kappa,11}\beta_u e_{\lambda,11}$. Then, an easy calculation shows that

$$(13) \qquad X_1, \cdots, X_t$$

is the K-basis of a vector space $\bar{V}$ which has $\bar{\mathfrak{a}}$ as left operator system. Then $\bar{V}$ is the representation space of a representation $\bar{\mathfrak{A}}$ of $\bar{\mathfrak{a}}$ corresponding to the representation $\mathfrak{A}$ of $\mathfrak{a}$. We observe that if $\bar{V}_\tau = KX_\tau + KX_{\tau+1} + \cdots + KX_t$ then

$$(14) \qquad \bar{V} = \bar{V}_1 \supset \bar{V}_2 \supset \cdots \supset \bar{V}_t \supset (0)$$

is a composition series for the $\bar{\mathfrak{a}}$-space $\bar{V}$.

Conversely, if we have a representation $\bar{\mathfrak{A}}$ of $\bar{\mathfrak{a}}$, we may assume that a basis adapted to a composition series of the representation space $\bar{V}$ of $\bar{\mathfrak{A}}$ has been chosen in the form (13) such that for each X_h there is some ν_h for which $X_h = e_{\nu_h,11}X_h$ $(h = 1, 2, \cdots, t)$. We form a basis for a representation space V of $\mathfrak{a}$ by adding to the basis of $\bar{V}$, vectors

$$(15) \qquad \begin{aligned} e_{\nu_h,a1}X_h \qquad a = 2, \cdots, f_\nu \quad \text{where} \quad & X_h = e_{\nu_h,11}X_h \\ & h = 1, 2, \cdots, t. \end{aligned}$$

and requiring the associativity condition that for $\alpha, \beta \in \mathfrak{a}$

$$(16) \qquad \alpha(\beta X_h) = (\alpha\beta)X_h.$$

We observe that this is just the inverse of the process of passing from $\mathfrak{a}$-spaces to $\bar{\mathfrak{a}}$-spaces.

In the above method of obtaining an $\bar{\mathfrak{a}}$-space $\bar{V}$ from the $\mathfrak{a}$-space V we used a particular composition series of V, and a particular choice of residues X_h. The methods used in 1.6 adapted to the case where $\mathfrak{a}$ is a left operator system only, shows that $\bar{V}$ is independent of the choice of composition series for V or of the residues X_h. Similar remarks hold for the inverse process of passing from an $\bar{\mathfrak{a}}$-space $\bar{V}$ to an $\mathfrak{a}$-space V. It follows that there is a (1-1) correspondence between the $\mathfrak{a}$-spaces and the $\bar{\mathfrak{a}}$-spaces, and consequently a (1-1) correspondence between representations of $\mathfrak{a}$ and those of $\bar{\mathfrak{a}}$.

An examination of the process of passing from the $\mathfrak{a}$-space V to the corre-

sponding $\bar{\mathfrak{a}}$-space $\bar{V}$, and of the inverse process, shows easily that the representations $\mathfrak{A}$ and $\bar{\mathfrak{A}}$ have corresponding irreducible constituents, corresponding indecomposable constituents, corresponding Loewy constituents,[9] etc. In fact, if we calculate $\mathfrak{A}$ by the basis (12), of V, we have the following scheme:

$$(17)\quad \alpha(X_{1,1}, \cdots, X_{1,f_{\nu_1}}; \cdots; X_{t,1}, \cdots, X_{t,f_{\nu_t}}) = (X_{1,1}, \cdots, X_{1,f_{\nu_1}}; \cdots; X_{t,1}, \cdots, X_{t,f_{\nu_t}}) \begin{bmatrix} A_{11}(\alpha) & & & \\ A_{21}(\alpha) & A_{22}(\alpha) & & \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ A_{t1}(\alpha) & A_{t2}(\alpha) & \cdots & A_{tt}(\alpha) \end{bmatrix}$$

while for $\bar{\mathfrak{A}}$ we have, using (13),

$$(18)\quad \bar{\alpha}(X_1, \cdots, X_t) = (X_1, \cdots, X_t) \begin{bmatrix} \bar{A}_{11}(\bar{\alpha}) & & & \\ \bar{A}_{21}(\bar{\alpha}) & \bar{A}_{22}(\bar{\alpha}) & & \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \bar{A}_{t1}(\bar{\alpha}) & \bar{A}_{t2}(\bar{\alpha}) & \cdots & \bar{A}_{tt}(\bar{\alpha}) \end{bmatrix}$$

If we calculate the matrix in $\mathfrak{A}$ which corresponds to $\bar{\alpha}$ we find that in each $A_{ij}(\bar{\alpha})$, the only coefficient different from zero is that in the upper left corner, and recalling that $X_{h,1} = X_h$ we see that this coefficient is $\bar{A}_{ij}(\bar{\alpha})$. Thus, to obtain $\bar{\mathfrak{A}}$ from A we may take the matrices of $\mathfrak{A}$ corresponding to the elements $\bar{\alpha}$ and 'deflate' them by striking out all rows and columns except those passing through the upper corners of the simple parts $\mathfrak{A}_{ij}$. The simple parts $\mathfrak{A}_{ij}(\bar{\mathfrak{A}}_{ij})$ are linear combinations of the elementary modules $\mathfrak{H}_u(\bar{\mathfrak{H}}_u)$, and if $\mathfrak{A}_{ij} = \sum_{u=1}^{r} d_{ij}^{u} \mathfrak{H}_u$, where the d_{ij}^{u} are fixed elements of K, then $\bar{\mathfrak{A}}_{ij} = \sum_{u=1}^{r} d_{ij}^{u} \bar{\mathfrak{H}}_u$.[10] This completes the discussion of 2.1.

3. Classes of algebras

An equivalence relation among algebras over K may now be set up as follows. We say that $\mathfrak{a}$ is *similar* to $\mathfrak{b}$ (and write $\mathfrak{a} \sim \mathfrak{b}$), if the basic algebra $\bar{\mathfrak{a}}$ of $\mathfrak{a}$ is isomorphic to the basic algebra $\bar{\mathfrak{b}}$ of $\mathfrak{b}$. It follows at once that the algebras over K are classified by this relation into disjoint classes of equivalent algebras.

2.1 shows that all algebras belonging to a particular class have corresponding representations. In fact, if $\mathfrak{a} \sim \mathfrak{b}$, we shall have correspondences of the following form: $\mathfrak{a}$-space $V \leftrightarrow \bar{\mathfrak{a}}$-space $\bar{V} \leftrightarrow \bar{\mathfrak{b}}$-space $\bar{W} \leftrightarrow \mathfrak{b}$-space W. The representations $\mathfrak{A}$ and $\mathfrak{B}$ obtained from V, W respectively, may be written with corresponding simple parts $\mathfrak{A}_{ij}$, $\mathfrak{B}_{ij}$, but these may differ in their dimensions (see, for example, $\mathfrak{A}$, $\bar{\mathfrak{A}}$ in (17) (18)).

[9] For a discussion of Loewy constituents, see [3], §5.

[10] This may be seen by computation of $\mathfrak{A}$ and $\bar{\mathfrak{A}}$ by means of (8), (17), (18). See also [12], Th. 4.

The following theorem indicates that we may define multiplication of classes by taking as the product of two classes, the class containing the direct product of any pair of representatives of the two classes:

3.1 *If* $\mathfrak{a} = \mathfrak{b} \times \mathfrak{c}$ *then* $\bar{\mathfrak{a}} = \bar{\mathfrak{b}} \times \bar{\mathfrak{c}}$.

PROOF: Let $\beta_1, \beta_2, \cdots, \beta_p$ be a system of primitive elements for $\mathfrak{b}$, and $\gamma_1, \cdots, \gamma_q$ a system of primitive elements for $\mathfrak{c}$. Then one may verify that $\beta_i\gamma_j$ $(i = 1, 2, \cdots, p; j = 1, 2, \cdots, q)$ is a system of primitive elements for $\mathfrak{b} \times \mathfrak{c}$. The algebra generated by these elements $\beta_i\gamma_j$ is $\bar{\mathfrak{b}} \times \bar{\mathfrak{c}}$.

4. Commutator algebras

Let V be an $\mathfrak{a}$-left space, and let

$$V = V_{11} + \cdots + V_{1s_1} + \cdots + V_{g1} + \cdots + V_{gs_g} \tag{19}$$

denote the decomposition of V into a direct sum of indecomposable $\mathfrak{a}$-spaces where $V_{\rho 1} \cong V_{\rho 2} \cong \cdots \cong V_{\rho s_\rho}$. Let $\mathfrak{a}'$ denote the algebra of $\mathfrak{a}$-homomorphisms of V; $\mathfrak{a}'$ is again an algebra with unit element over K. We consider $\mathfrak{a}'$ also as a left operator system of V.[11] If $\alpha \in \mathfrak{a}$, $\alpha' \in \mathfrak{a}'$, $X \in V$ then

$$\alpha'(\alpha X) = \alpha(\alpha' X). \tag{20}$$

In terms of representations we have that if V as an $\mathfrak{a}$-space produces the representation $\mathfrak{A}$ of $\mathfrak{a}$, and as an $\mathfrak{a}'$-space produces the representation $\mathfrak{A}'$ of $\mathfrak{a}'$, then $\mathfrak{A}'$ is the commutator algebra of $\mathfrak{A}$. Here $\mathfrak{A}$ need not be a faithful representation of $\mathfrak{a}$, but $\mathfrak{A}'$ is faithful for $\mathfrak{a}'$. It is to be noted that $\mathfrak{a}'$ depends on the space V, and, as we shall see below, its structure may vary considerably with different choices of V.

The main theorem on homomorphism algebras (or what is equivalent, commutator algebras) is

4.1 *The* $\mathfrak{a}$*-homomorphism algebra* $\mathfrak{a}'$ *of an* $\mathfrak{a}$*-space* V *of form* (19) *may be written as the direct sum*

$$\mathfrak{a}' = \tilde{\mathfrak{a}}_1 + \cdots + \tilde{\mathfrak{a}}_g + \mathfrak{n}' \tag{21}$$

where $\tilde{\mathfrak{a}}_\rho$ *denotes a simple algebra of* $\mathfrak{a}$*-homomorphisms of* $V_\rho = V_{\rho 1} + \cdots + V_{\rho}s_\rho$, *and* $\mathfrak{n}'$ *is the radical of* $\mathfrak{a}'$.[12]

PROOF: Denote by $e'_{\rho,p1}$ an element of $\mathfrak{a}'$ which maps $V_{\rho 1}$ onto $V_{\rho p}$ isomorphically, and maps $V_{\tau a}$, where the pair $(\tau, a) \neq$ the pair $(\rho, {}_1)$, onto zero. In particular, let $e'_{\rho,11}$ $(\rho = 1, 2, \cdots, g)$ effect the identity mapping of $V_{\rho 1}$ onto itself. Let $e'_{\rho,1p}$ be an element of $\mathfrak{a}'$ which maps $V_{\tau a}$ onto zero for $(\tau, a) \neq (\rho, p)$ and $V_{\rho p}$ onto $V_{\rho 1}$ in such a way that $e'_{\rho,1p}e'_{\rho,p1} = e'_{\rho,11}$. Then $e'_{\rho,pq} = e'_{\rho,p1}e'_{\rho,1q}$

[11] In place of this, E. Artin strongly favors using the inverse algebra of $\mathfrak{a}'$ as a right operator system, in which case the commutativity relations (20) is replaced by an associativity relation (see (5)). We choose, however, in this section to take $\mathfrak{a}'$ as left operator system, as this seems a little more directly related to the idea of commutator algebra of a representation.

[12] The material of this section follows rather directly from ring theory established by other writers. See, for example, Brauer [2], lectures on the theory of rings by E. Artin (forthcoming), and the reference in footnote 2.

maps $V_{\rho q}$ onto $V_{\rho p}$, and maps $V_{\tau a}$ onto zero, $(\tau, a) \neq (\rho, q)$. Further, $e'_{\rho,pq}e'_{\sigma,uv} = \delta_{\rho\sigma}\delta_{qu}e'_{\rho,pv}$ (δ_{ab} the identity mapping for $a = b$, and the 0-mapping for $a \neq b$).

We observe that each $e'_{\rho,pp}$ is a primitive idempotent of $\mathfrak{a}'$. For, if $e'_{\rho,pp} = e'_1 + e'_2$, where e'_1, e'_2 are mutually orthogonal idempotents $\neq 0$ of $\mathfrak{a}'$, then $V_{\rho p} = e'_{\rho,pp}V_{\rho p} = e'_1V_{\rho p} + e'_2V_{\rho p}$ would be the direct sum of the $\mathfrak{a}$-spaces $e'_1V_{\rho p}$, $e'_2V_{\rho p}$, contrary to the assumption that $V_{\rho p}$ is indecomposable. We next observe that $\bar{\mathfrak{a}}_\rho = \sum_{p,q=1}^{s} Ke'_{\rho,pq}$ is a simple algebra of degree s_ρ^2 over K. Also $\sum_\rho \sum_{p=1}^{s} e'_{\rho,pp}$ is a decomposition of the unit element e' of $\mathfrak{a}'$ into a sum of mutually orthogonal primitive idempotents. It follows that $\bar{\mathfrak{a}}_1 + \cdots + \bar{\mathfrak{a}}_g$ is a semisimple subalgebra of $\mathfrak{a}'$ which is isomorphic to $\mathfrak{a}'/\mathfrak{n}'$ and (21) results.

4.2 *Let* $\mathfrak{a}'_1 \supset \mathfrak{a}'_2 \supset \cdots \supset \mathfrak{a}'_y \supset (0)$ *denote a composition series of* $\mathfrak{a}'$ *considered as an* $(\mathfrak{a}', \mathfrak{a}')$ *space, and let* $\mathfrak{a}'_v/\mathfrak{a}'_{v+1}$ *be a composition factor group of type* (ρ, σ), $\rho, \sigma = 1, 2, \cdots$, *or* g. *Then there is an element* γ'_v *in* $\mathfrak{a}'_v$ *such that* (1) γ'_v *maps* $V_{\sigma 1}$ *into* $V_{\rho 1}$ *and maps all other* $V_{\tau a}$ *into* 0, (2) $\mathfrak{a}'_v/\mathfrak{a}'_{v+1} = \bar{\mathfrak{a}}_\rho\langle\gamma'_v\rangle\bar{\mathfrak{a}}_\sigma$.

PROOF: This is an immediate consequence of 1.4 applied to the irreducible $(\mathfrak{a}', \mathfrak{a}')$-space $\mathfrak{a}'_v/\mathfrak{a}'_{v+1}$. We have from 1.4 that $\gamma'_v \in \mathfrak{a}'_v$ exists such that $e'_{\rho,11}\gamma'_ve'_{\sigma,11} = \gamma'_v$ and that (2) holds. But the homomorphism $e'_{\rho,11}\gamma'_ve'_{\sigma,11}$ maps $V_{\sigma 1}$ into $V_{\rho 1}$, and $V_{\tau a}$ onto zero if $(\tau, {}_a) \neq (\sigma, {}_1)$, so that (1) also is satisfied.

By applying 1.5 we may obtain a Cartan basis for $\mathfrak{a}'$, and by means of its elements define a system of elementary modules of $\mathfrak{a}'$.

We call the $\mathfrak{a}$-space $\hat{V}$ formed by taking one $V_{\rho p}$ from each set of isomorphic indecomposable subspaces of V, the *reduced space* of V. We may assume

$$\hat{V} = V_{11} + \cdots + V_{g1}.$$

4.3 *The basic algebra of the* $\mathfrak{a}$*-homomorphism algebra* $\mathfrak{a}'$ *of the* $\mathfrak{a}$*-space* V *is the* $\mathfrak{a}$*-homomorphism algebra of the reduced space* $\hat{V}$ *of* V.

PROOF: The primitive elements γ'_v, $v = 1, 2, \cdots, g$, (cf. 4.2) generate the basic algebra $\bar{\mathfrak{a}}'$ of $\mathfrak{a}'$. It follows from (1) in 4.2 that each element γ'_v gives an $\mathfrak{a}$-homomorphism of $\hat{V}$. We have only to verify that every $\mathfrak{a}$-homomorphism of $\hat{V}$ is linearly dependent on the γ'_v. $\bar{e}' = \sum_{p=1}^{g} e'_{\rho,11}$ is the identity homomorphism of $\hat{V}$. Then any homomorphism θ of $\hat{V}$ satisfies $\theta = \bar{e}'\theta\bar{e}' = \sum_{\rho,\sigma} e'_{\rho,11}\theta e'_{\sigma,11}$. θ may be extended to a homomorphism of V by setting $\theta \cdot V_{\tau a} = (0)$ for $a \neq 1$. As a homomorphism of V, $\theta = \sum_{\rho,\sigma=1}^{g} e'_{\rho,11}\theta e'_{\sigma,11}$ is linearly dependent on the primitive elements γ'_v.

Theorem 4.1 provides some information concerning the irreducible constituents of the commutator algebra $\mathfrak{A}'$. Since $\mathfrak{A}'$ is a faithful representation of $\mathfrak{a}'$, there appears in $\mathfrak{A}'$, corresponding to the simple algebra $\bar{\mathfrak{a}}_\rho$ in (21), an irreducible constituent $\mathfrak{F}'_\rho$ of degree s_ρ ($\rho = 1, 2, \cdots, g$). It has been shown elsewhere that the multiplicity of $\mathfrak{F}'_\rho$ equals the K-dimension of $V_{\rho 1}$.[12] Thus *to each distinct indecomposable part* $\mathfrak{U}_\rho$ *of* $\mathfrak{A}$ *there corresponds an irreducible part* $\mathfrak{F}'_\rho$ *of* $\mathfrak{A}'$ *such that the degree of* $\mathfrak{F}'_\rho$ *is the multiplicity of* $\mathfrak{U}_\rho$ *in* $\mathfrak{A}$, *and the multiplicity of* $\mathfrak{F}'_\rho$ *in* $\mathfrak{A}'$ *is the degree of* $\mathfrak{U}_\rho$.

For $\mathfrak{a}$ simple, and hence also for $\mathfrak{a}$ semisimple, the commutator algebras of faithful representations of $\mathfrak{a}$ are all similar in the sense of §3. However, this

may not be true if $\mathfrak{a}$ has a radical, as the following example shows. Let $\mathfrak{a}$ be an algebra over K having $\mathfrak{A} = \begin{pmatrix} \mathfrak{H}_1 & 0 \\ \mathfrak{H}_{21} & \mathfrak{H}_2 \end{pmatrix}$ and $\mathfrak{B} = \begin{pmatrix} \mathfrak{H}_1 & 0 & 0 \\ \mathfrak{H}_{21} & \mathfrak{H}_2 & 0 \\ 0 & 0 & \mathfrak{H}_1 \end{pmatrix}$ as faithful representations, where $\mathfrak{H}_1$, $\mathfrak{H}_2$, $\mathfrak{H}_{21}$ denote elementary modules of 1×1 matrices. The commutator algebra of $\mathfrak{A}$ is of form $\mathfrak{C} = \begin{pmatrix} \mathfrak{M}_1 & 0 \\ 0 & \mathfrak{M}_1 \end{pmatrix}$, and for $\mathfrak{B}'$ is $\mathfrak{C}' = \begin{pmatrix} \mathfrak{M}_1 & 0 & 0 \\ 0 & \mathfrak{M}_1 & 0 \\ \mathfrak{M}_3 & 0 & \mathfrak{M}_2 \end{pmatrix}$ where the $\mathfrak{M}_i$ are elementary modules consisting of 1×1 matrices, and $\mathfrak{M}_2 \neq \mathfrak{M}_1$. Here $\mathfrak{C}$ is not similar to $\mathfrak{C}'$.

We may obtain, however, the following theorem which is almost obvious from the results derived in §§2, 3.

4.4 *If $\mathfrak{a}$ is similar to $\mathfrak{b}$, and the $\mathfrak{a}$-space V corresponds to the $\mathfrak{b}$-space W then the $\mathfrak{a}$-homomorphism algebra $\mathfrak{a}'$ of V is similar to the $\mathfrak{b}$-homomorphism algebra $\mathfrak{b}'$ of W.*

Proof: It is sufficient to show that the theorem holds for the $\mathfrak{a}$-space V and the corresponding $\bar{\mathfrak{a}}$-space $\bar{V}$. We consider the reduced spaces $\hat{V}$, $\hat{\bar{V}}$ (cf. 4.3). Let

$$\hat{V} = V_1 + V_2 + \cdots + V_g ; \qquad \hat{\bar{V}} = \bar{V}_1 + \bar{V}_2 + \cdots + \bar{V}_g$$

be decompositions of $\hat{V}$, $\hat{\bar{V}}$ into direct sums of distinct indecomposable subspaces. Let e' be the identity homomorphism of $\hat{V}$, and e'_ρ the $\mathfrak{a}$-homomorphism which maps V_ρ identically on itself, and V_λ on (0), $\lambda \neq \rho$. Then for any $\alpha' \in \mathfrak{a}'$, $\alpha' = e'\alpha'e' = \sum_{\rho,\sigma} e'_\rho\alpha'e'_\sigma$, where $\alpha'_{\rho\sigma} = e'_\rho\alpha'e'_\sigma$ maps V_σ into V_ρ, and V_τ on (0), $\tau \neq \sigma$. We take composition series

$$V_\theta = V_1^\theta \supset V_2^\theta \supset \cdots \supset V_{t_\theta}^\theta ;$$

and apply the method of §2 to obtain bases of the form (12) for V_σ and V_ρ namely

$$X_{h,a}^\theta = e_{\nu_h,a1}X_h^\theta \qquad \begin{matrix} h = 1, 2, \cdots, t_\theta \\ a = 1, 2, \cdots, f_{\nu_h} \end{matrix} \qquad (\theta = \rho, \sigma)$$

Here the $X_h^\theta = e_{\nu_h,11}X_h^\theta$, $h = 1, 2, \cdots, t_\theta$ form a basis for $\bar{V}_\theta$. If now, under $\alpha'_{\rho\sigma}$, $X_h^\sigma \to Y_h^\rho \in V_\rho$, then

(22)
$$e_{\nu_h,a1}X_h^\sigma \to e_{\nu_h,a1}Y_h^\rho$$

in particular,

$$X_h^\sigma \to e_{\nu_h,11}Y_h^\rho$$

which shows that $Y_h^\rho \in \bar{V}_\rho$. Let us denote the $\bar{\mathfrak{a}}$-mapping of $\hat{\bar{V}}$ which takes X_h^σ into Y_h^ρ $(h = 1, 2, \cdots, t_\sigma)$ by $\bar{\alpha}'_{\rho\sigma}$. Then $\bar{\alpha}' = \sum \bar{\alpha}'_{\rho\sigma}$ is an $\bar{\mathfrak{a}}$-mapping of $\hat{\bar{V}}$ determined by the $\mathfrak{a}$-mapping α' of $\hat{V}$. Inversely, we may start with an $\bar{\mathfrak{a}}$-mapping of $\hat{\bar{V}}$, and use the relations (22) to obtain an $\mathfrak{a}$-mapping of $\hat{V}$. It is then

easy to verify that the $\mathfrak{a}$-homomorphism algebra of $\hat{V}$ is isomorphic to the $\bar{\mathfrak{a}}$-homomorphism algebra of $\hat{\bar{V}}$, and from 4.3 it then follows that the $\mathfrak{a}$-homomorphism algebra of V is similar to the $\mathfrak{a}$-homomorphism algebra of $\bar{V}$.

II. Symmetric Functions of $\bar{\mathfrak{a}}$

5. Preliminaries

Brauer, Nakayama, and one of the authors have discussed in [5], [11], [9], [10], a class of algebras which they called symmetric algebras. An algebra is called symmetric if in $\mathfrak{a}$, considered as a vector space, there exists a hyperplane which contains all commutator elements $\alpha\beta - \beta\alpha$ of $\mathfrak{a}$ but does not contain a one-sided $\mathfrak{a}$-ideal other than (0). Here we shall discuss hyperplanes of $\mathfrak{a}$ which contain all commutator elements, such a hyperplane we shall call a symmetric hyperplane. However, we propose a slight change of language. We shall say that a function of $\mathfrak{a}$ is linear symmetric, if in addition to linearity, $\varphi(\alpha\beta) = \varphi(\beta\alpha)$ for all $\alpha, \beta \epsilon \mathfrak{a}$. A symmetric hyperplane yields a linear symmetric function of $\mathfrak{a}$, and conversely. For if the symmetric hyperplane h consists of all elements $\alpha \epsilon \mathfrak{a}$ such that $\psi(\alpha) = 0$, ψ a linear function, then the condition that all commutators $\alpha\beta - \beta\alpha$ appear in h implies that $\psi(\alpha\beta) = \psi(\beta\alpha)$. In our discussion it is more convenient, generally, to speak of linear symmetric functions of $\mathfrak{a}$ rather than of the corresponding symmetric hyperplanes. For brevity, let us now speak of symmetric functions, it being understood that these shall be linear.

The characters of the representations of $\mathfrak{a}$ form an important class of symmetric functions of $\mathfrak{a}$.

We shall have occasion to refer to the regular representations of $\mathfrak{a}$ which we now define. Let $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ be a basis for $\mathfrak{a}$. For every α in $\mathfrak{a}$ we have equations

$$\epsilon_k\alpha = \sum_k r_{kl}(\alpha)\, \epsilon_l$$

$$\alpha\, \epsilon_k = \sum_m s_{mk}(\alpha)\, \epsilon_m$$

where the coefficients $r_{kl}(\alpha)$, $s_{mk}(\alpha)$ lie in K. The first regular representation, $\mathfrak{R}$ of $\mathfrak{a}$ is given by the homomorphism $\alpha \rightarrow R(\alpha) = (r_{ij}(\alpha))$ and the second regular representation, $\mathfrak{S}$ of $\mathfrak{a}$, by $\alpha \rightarrow S(\alpha) = (s_{ij}(\alpha))$ where i, j denote row and column index respectively. It has been shown that an algebra is symmetric if and only if there exists a symmetric non-singular matrix T which transforms $\mathfrak{R}$ into $\mathfrak{S}$.[13]

6. Symmetric functions of $\mathfrak{a}$ and of $\bar{\mathfrak{a}}$

Here we shall prove

6.1. *There is a* (1-1) *correspondence between the symmetric functions of* $\mathfrak{a}$ *and those of the basic algebra* $\bar{\mathfrak{a}}$ *of* $\mathfrak{a}$.

Proof: Let φ be a symmetric function of $\mathfrak{a}$, then φ maps all commutators

[13] [11], p. 654.

$\alpha\beta$-$\beta\alpha$ of $\mathfrak{a}$ on zero. We use here as basis of $\mathfrak{a}$ the Cartan basis system (7). A commutator determined by two of these basis elements is of form.

$$\gamma = e_{\kappa_u,a1}\beta_u e_{\lambda_u,1b} e_{\kappa_v,c1}\beta_v e_{\lambda_v,1d} - e_{\kappa_v,c1}\beta_v e_{\lambda_v,1d} e_{\kappa_u,a1}\beta_u e_{\lambda_u,1b} .$$

We distinguish three cases:

1) Neither of the pairs $\begin{cases}\lambda_u = \kappa_v \\ b = c\end{cases}, \begin{cases}\lambda_v = \kappa_u \\ d = a\end{cases}$ is satisfied. Then $\gamma = 0$.

2) One of the pairs, say the first, is satisfied, and the other is not. Then $\gamma = e_{\kappa_u,a1}\beta_u\beta_v e_{\lambda_v,1d}$.

3) Both pairs are satisfied. Then

$$\gamma = e_{\kappa_u,a1}\beta_u\beta_v e_{\kappa_u,1a} - e_{\kappa_v,b1}\beta_v\beta_u e_{\kappa_v,1b} .$$

If in case 2) we take $e_{\kappa_v,c1}\beta_v e_{\lambda_v,1d}$ to be the element $e_{\lambda_u,bd}$ we obtain $\gamma = e_{\kappa_u,a1}\beta_u e_{\lambda_u,1d}$ if $a \neq d$ or $\kappa_u \neq \lambda_u$ hence

$$\varphi(e_{\kappa_u,a1}\beta_u e_{\lambda_u,1d}) = 0, \quad \text{if} \quad a \neq d \quad \text{or} \quad \kappa_u \neq \lambda_u . \tag{23}$$

As a special case of 3) we take β_u to be of unmixed type (κ_u, κ_u)[14] and $e_{\kappa_v,c1}\beta_v e_{\lambda_v,1d}$ to be $e_{\kappa_u,ba}$, and obtain $\gamma = e_{\kappa_u,a1}\beta_u e_{\kappa_u,1a} - e_{\kappa_u,b1}\beta_u e_{\kappa_u,1b}$. Thus for β_u of unmixed type

$$\varphi(e_{\kappa_u,a1}\beta_u e_{\kappa_u,1a}) = \varphi(e_{\kappa_u,b1}\beta_u e_{\kappa_u,1b}). \tag{24}$$

Let us denote by c_{uvw} the multiplication constants of the elements $\beta_u (u = 1, 2, \cdots, r)$, that is,

$$\beta_u\beta_v = \sum_{w=1}^{r} c_{uvw}\beta_w . \tag{25}$$

We shall now define a function $\bar{\varphi}$ of $\bar{\mathfrak{a}}$ which, in addition to being linear, satisfies the relations

$$\begin{aligned} &\bar{\varphi}(\beta_w) = 0 \quad \text{if} \quad \beta_w \text{ of mixed type}^{14} \\ &\bar{\varphi}(\beta_w) = \varphi(e_{\kappa_w,a1}\beta_w e_{\kappa_w,1a}) \quad \text{if} \quad \beta_w \text{ of unmixed type,} \end{aligned} \tag{26}$$

the latter relation according to (24) being independent of a. Then since $\varphi(\gamma) = 0$, for γ as in case 3) we obtain successively:

$$\begin{aligned} &\varphi(e_{\kappa_u,a1}\beta_u\beta_v e_{\kappa_u,1a}) = \varphi(e_{\kappa_v,b1}\beta_v\beta_u e_{\kappa_v,1b}), \\ &\varphi(\sum_w c_{uvw} e_{\kappa_u,a1}\beta_w e_{\kappa_u,1a}) = \varphi(\sum_y c_{vuy} e_{\kappa_v,b1}\beta_y e_{\kappa_v,1b}), \\ &\bar{\varphi}(\sum_w c_{uvw}\beta_w) = \bar{\varphi}(\sum_y c_{vuy}\beta_y) \\ &\bar{\varphi}(\beta_u\beta_v) = \bar{\varphi}(\beta_v\beta_u). \end{aligned}$$

---

[14] A quantity of type (κ, λ) is of mixed type if $\kappa \neq \lambda$, and is of unmixed type if $\kappa = \lambda$.

Here $\beta_u\beta_v$ and $\beta_v\beta_u$ are of unmixed types (κ_u, κ_u), (κ_v, κ_v) respectively. If, however, $\beta_u\beta_v \neq 0$ is of mixed type, then $\beta_v\beta_u = 0$, and the first relation in (26) shows that in this case also $\bar{\varphi}(\beta_u\beta_v) = \bar{\varphi}(\beta_v\beta_u)$. This shows that the symmetric function φ of $\mathfrak{a}$ determines a symmetric function $\bar{\varphi}$ of $\bar{\mathfrak{a}}$. The inverse process is now evident, and the proof of 6.1 is completed.

Let us call the function ψ_w of $\mathfrak{a}$ defined by

$$\psi_w(\alpha) = \text{trace } H_w(\alpha),$$

where $\mathfrak{H}_w$ is an elementary module of unmixed type, the character of $\mathfrak{H}_w$. Then we may show

6.2. *Every symmetric function of $\mathfrak{a}$ is expressible as a linear combination of the characters of the elementary modules of unmixed type.*

PROOF: If φ is a symmetric function of $\mathfrak{a}$, then for $\alpha \in \mathfrak{a}$

$$\varphi(\alpha) = \varphi(\sum_{a,b,u} h^u_{ab}(\alpha)e_{\kappa_u,a1}\beta_u e_{\lambda_u,1b}),$$

and by (23) this gives

$$\varphi(\alpha) = \varphi(\sum_{a,w} h^w_{aa}(\alpha)e_{\kappa_w,a1}\beta_w e_{\kappa_w,1a})$$

where now the range of summation for w is determined by the elements β_w of unmixed type. It follows from (24) and (26) that

$$\varphi(\alpha) = \sum_w \bar{\varphi}(\beta_w)\psi_w(\alpha), \tag{27}$$

and 6.2 is proved.

If $\bar{\varphi}$ is a symmetric function of $\bar{\mathfrak{a}}$, then the symmetric function of $\mathfrak{a}$, which according to 6.1 is determined by $\bar{\varphi}$, is $\sum_w \bar{\varphi}(\beta_w)\psi_w$.

7. Center elements and symmetric functions

From 6.2 it follows that the number q of linearly independent symmetric functions of $\mathfrak{a}$ is at most equal to the number $s = \sum_{\kappa=1}^{k} c_{\kappa\kappa}$ of elementary modules of unmixed type. If $q = s$ we say that the algebra $\mathfrak{a}$ is *completely permutative.* For a completely permutative algebra $\mathfrak{a}$ the character of each elementary module of unmixed type is a symmetric function of $\mathfrak{a}$.

Let us denote by p the rank of the center of $\mathfrak{a}$. Since for an element ζ of the center $H_u(\zeta) = 0$, $\mathfrak{H}_u$ of mixed type; $H_w(\zeta) = c_w(\zeta)\delta_{ij}$ where $c_w(\zeta) \in K$, for $\mathfrak{H}_w$ of unmixed type, then $p \leqq s$.[15] For a completely permutative algebra $p \leqq q = s$. The inequality may hold. For example, take

$$\mathfrak{A} = \begin{pmatrix} \mathfrak{H}_1 & \\ \mathfrak{H}_3 & \mathfrak{H}_2 \end{pmatrix}$$

where the $\mathfrak{H}_i$ denote elementary modules, such that $\mathfrak{H}_1$, $\mathfrak{H}_2$ are distinct irreducible constituents of $\mathfrak{A}$, and $\mathfrak{H}_3$ belongs to the radical. Then the characters of $\mathfrak{H}_1$, $\mathfrak{H}_2$ are symmetric functions of $\mathfrak{a}$, and $q = s = 2$, while $p = 1$.

---

[15] [12], Th. 9.

One may, on the other hand, construct cases where q is less than p. Consider

$$\mathfrak{A} = \begin{pmatrix} \mathfrak{H}_1 & & & \\ \mathfrak{H}_3 & \mathfrak{H}_2 & & \\ \mathfrak{H}_5 & \mathfrak{H}_4 & \mathfrak{H}_1 & \\ \mathfrak{H}_6 & \mathfrak{H}_7 & 0 & \mathfrak{H}_1 \end{pmatrix}$$

where again the $\mathfrak{H}_i$ denote distinct elementary modules. Here $s = 4$. Three linearly independent elements α, β, γ of the center are obtained by taking:

for α, $H_1(\alpha)$, $H_2(\alpha)$ to be unit matrices, and $H_i(\alpha) = (0)$, $i \neq 1, 2$;

for β, $H_5(\beta)$ to be a unit matrix, and $H_i(\beta) = (0)$, $i \neq 5$;

for γ, $H_6(\gamma)$ to be a unit matrix, and $H_i(\gamma) = (0)$, $i \neq 6$.

Any symmetric function of $\mathfrak{a}$ is a linear combination of the characters of $\mathfrak{H}_1$, $\mathfrak{H}_2$, $\mathfrak{H}_5$, and $\mathfrak{H}_6$. The characters of $\mathfrak{H}_5$, and $\mathfrak{H}_6$ are unsymmetric, and it follows that $q = 2$.

Nakayama[16] has shown that for a symmetric algebra there exists a (1-1) correspondence between symmetric hyperplanes and elements of the center, and hence $p = q$. In fact, let $T = (t_{ij}) = T'$ be a symmetric matrix which transforms the first regular representation, R, into second regular representation, $\mathfrak{S}$, that is, $\mathfrak{S}T = T\mathfrak{R}$, $|T| \neq 0$. Let $\epsilon_1, \cdots, \epsilon_n$ again denote a basis of $\mathfrak{a}$, and set $\bar{\epsilon}_i = \sum_j t_{ij}\epsilon_j$. If φ is a symmetric function of $\mathfrak{a}$, then $\sum_{i=1}^n \varphi(\bar{\epsilon}_i)\epsilon_i$ is an element of the center of $\mathfrak{a}$. Conversely, if $\sum_{j=1}^n \theta(\epsilon_j)\epsilon_j$ is an element of the center of a symmetric algebra $\mathfrak{a}$, then a symmetric function φ of $\mathfrak{a}$ may be defined by $\varphi(\bar{\epsilon}_j) = \theta(\epsilon_j)$. For the first of these statements it is not necessary that the symmetric matrix T be non-singular.

8. The case $p = s$

We shall show,

8.1 *In order that the rank p of the center of $\mathfrak{a}$ should equal $s = \sum_{\kappa=1}^k c_{\kappa\kappa}$ it is necessary and sufficient that a) there be no elementary modules of $\mathfrak{a}$ of mixed type and b) $\mathfrak{a}$ be completely permutative, $q = s$.*

Proof: i) An element ζ of the center is of form $\zeta = \sum_w c_w(\zeta)e^{(w)}$ where $e^{(w)} = \sum_a e_{\kappa_w,a1}\beta_w e_{\kappa_w,1a}$, β_w of unmixed type (κ_w, κ_w).[15] If the number p of linearly independent center elements equals s, then each of the s elements $e^{(w)}$ is an element of the center. In particular, $e_\kappa = \sum_a e_{\kappa,aa}$ is an element of the center. If now there were a basis element β_u of mixed type (κ, λ), we would have $e_\kappa\beta_u = \beta_u$, while $\beta_u e_\kappa = 0$. Thus, if $p = s$, there are no elementary modules of $\mathfrak{a}$ of mixed type.

It follows easily that the basic algebra $\bar{\mathfrak{a}}$ is commutative. For if $e^{(w)}$, $e^{(v)}$ are both of type (κ, κ), we have, since these elements are in the center.

$$e^{(w)} \cdot e^{(v)} = e^{(v)} \cdot e^{(w)}$$

$$\sum_a e_{\kappa,a1}\beta_w\beta_v e_{\kappa,1a} = \sum_b e_{\kappa,b1}\beta_v\beta_w e_{\kappa,1b}$$

---

[16] Cf. [9].

and multiplying both members on left and right by $e_{\kappa,1c}$, we obtain

$$\beta_w\beta_y = \beta_y\beta_w . \tag{28}$$

(28) evidently holds also when β_w, β_y have different types say (κ, κ) and (λ, λ), and so $\bar{\mathfrak{a}}$ is commutative.

Since $\bar{\mathfrak{a}}$ is commutative every linear function of $\bar{\mathfrak{a}}$ is symmetric. In particular, we obtain s linearly independent symmetric functions $\bar{\varphi}_w$ by setting

$$\bar{\varphi}_w(\beta_y) = 0, \qquad y \neq w; \qquad \bar{\varphi}_w(\beta_w) = 1 \qquad (w = 1, 2, \cdots, s). \tag{29}$$

Then 6.1 shows that there exist s linearly independent symmetric functions of $\mathfrak{a}$, so that $\mathfrak{a}$ is completely permutative.

ii) If there are no elementary modules of mixed type, then all Cartan basis elements β_w are of unmixed type and the basic algebra $\bar{\mathfrak{a}}$ is of order s. If further $\bar{\mathfrak{a}}$ is completely permutative, then it follows from 6.1 that there exist s linearly independent symmetric functions of $\bar{\mathfrak{a}}$. Hence every linear function of $\bar{\mathfrak{a}}$ is symmetric (in particular, the functions (29)) which implies that $\bar{\mathfrak{a}}$ is commutative. Then by reversing the argument in i) we find that the rank p of the center of $\mathfrak{a}$ is s.

From the above proof it is evident that the theorem might have been stated in the form: *The rank p of the center of $\mathfrak{a}$ is equal to $s = \sum_{\kappa=1}^{k} c_{\kappa\kappa}$ if and only if the basic algebra $\bar{\mathfrak{a}}$ of $\mathfrak{a}$ is commutative.*

9. Blocks

In foregoing papers, irreducible representations, characters, Cartan basis elements and other entities have been classified according to "blocks".[17] These blocks correspond to invariant subalgebras which are direct summands of $\mathfrak{a}$, and we shall identify the "blocks" with these summands. Let $\mathfrak{a}^{(\tau)}$ be such a block, and let $s_\tau = \sum_\lambda' c_{\lambda\lambda}$ where the summation extends over Cartan invariants associated with $\mathfrak{a}^{(\tau)}$. Assume now that the rank p_τ of the center of $\mathfrak{a}^{(\tau)}$ is equal to s_τ. By 8.1, $\mathfrak{a}^{(\tau)}$ has Cartan basis elements of unmixed type only. If $\mathfrak{a}^{(\tau)}$ contains elements of type (κ, κ), then all elements of $\mathfrak{a}^{(\tau)}$ are of type (κ, κ), since the elements of type (κ, κ) form a direct summand of $\mathfrak{a}^{(\tau)}$. We have proved

9.1 *If the center of a block $\mathfrak{a}^{(\tau)}$ has rank $s_\tau = \sum_\lambda' c_{\lambda\lambda}$, then all elements of $\mathfrak{a}^{(\tau)}$ are of the same unmixed type. Moreover, $\mathfrak{a}^{(\tau)}$ is completely permutative.*

Let us now take for a the group ring Γ of a finite group G formed with respect to a suitable modular field $\bar{K}$.[18] Let $\Gamma^{(\tau)}$ now be a block of $\bar{\Gamma}$, and let $F_1, \cdots, F_{y_\tau}$, and $Z_1, \cdots, Z_{x_\tau}$ denote the modular and the ordinary irreducible representations of G belonging to $\Gamma^{(\tau)}$. It may be shown that the rank p_τ of the center of $\Gamma^{(\tau)}$ is equal to x_τ.[19] Each ordinary representation Z may be taken as a modular representation $\bar{Z}$. We shall show:

---

[17] See [6], §9; [12], §6.

[18] Cf. [6].

[19] See, for instance, the discussion of blocks by R. M. Thrall, On the decomposition of modular tensors II, (forthcoming).

9.2. *The block $\Gamma^{(\tau)}$ of the group ring $\bar{\Gamma}$ is completely permutative if and only if all the ordinary representations Z_i belonging to $\Gamma^{(\tau)}$ remain irreducible when taken as modular representations.*

PROOF: *Only if.* Let q_τ denote the number of linearly independent symmetric functions of $\Gamma^{(\tau)}$. $\Gamma^{(\tau)}$ is a symmetric algebra,[20] so $p_\tau = q_\tau$, and $q_\tau = s_\tau$ by hypothesis. From 9.1 it follows that $\Gamma^{(\tau)}$ has elements of just one type, say (κ, κ), and $s_\tau = c_{\kappa\kappa}$. Then $\Gamma^{(\tau)}$ has only one irreducible representation, which we denote by F_κ. Further, $\bar{Z}_i$ has F_κ as its only irreducible constituent, $\bar{Z}_i \leftrightarrow d_{i\kappa}F_\kappa$, where the notation is to indicate that F_κ appears $d_{i\kappa}$ times as constituent of $\bar{Z}_i$. The relations

$$p_\tau = x_\tau = c_{\kappa\kappa} = \sum_{i=1}^{x_\tau} d_{i\kappa}^2 \tag{30}$$

[21]

show that $d_{i\kappa} = 1, i = 1, 2, \cdots, x_\tau$, that is, $\bar{Z}_i$ is irreducible.

If. Any two ordinary irreducible representations Z_i, Z_j of $\Gamma^{(\tau)}$ may be connected by a chain $Z_i, Z_a, \cdots, Z_b, Z_j$ such that neighboring members have at least one modular irreducible constituent in common. It follows that if the Z_i are each irreducible as modular representations, then there is just one modular irreducible representation F_κ of $\Gamma^{(\tau)}$, and $\bar{Z}_i = F_\kappa$, $d_{i\kappa} = 1$, $i = 1, 2, \cdots, x_\tau$. Then (30) holds, and $\Gamma^{(\tau)}$ is completely permutative.

10. Symmetric algebras

In case $\mathfrak{a}$ is a symmetric algebra, the question arises as to whether the basic algebra $\bar{\mathfrak{a}}$ is also symmetric. In terms of symmetric functions an algebra $\mathfrak{a}$ is symmetric if there exists a symmetric function of $\mathfrak{a}$ which does not map any right ideal of $\mathfrak{a}$ on zero other than the 0-ideal.

10.1. *An algebra $\mathfrak{a}$ is symmetric if and only if its basic algebra $\bar{\mathfrak{a}}$ is symmetric.*

PROOF: Let us suppose $\bar{\mathfrak{a}}$ is symmetric and that $\bar{\varphi}$ is a symmetric function of $\bar{\mathfrak{a}}$ which does not map any proper right ideal of $\bar{\mathfrak{a}}$ on 0. Let φ be the corresponding symmetric function of $\mathfrak{a}$, and assume that the principal right ideal $\mathfrak{r}$ generated by

$$\alpha = \sum_{u,ab} h_{ab}^u(\alpha)e_{\kappa_u,a1}\beta_u e_{\lambda_u,1b}$$

is mapped on 0 by φ, and that $h_{cd}^v(\alpha)e_{\kappa_v,c1}\beta_v e_{\lambda_v,1d} \neq 0$. Since $\zeta = \alpha \cdot e_{\lambda_v,d1}\beta_y e_{\kappa_v,1c} \epsilon \mathfrak{r}$, we have from (23), (26)

$$\begin{aligned}\varphi(\zeta) &= \varphi\Big(\sum_x h_{cd}^x(\alpha)e_{\kappa_v,c1}\beta_x\beta_y e_{\kappa_v,1c}\Big) \\ &= \bar{\varphi}\Big(\sum_x h_{cd}^x(\alpha)\beta_x\beta_y\Big) = 0\end{aligned}$$

where now the summation runs through the β_x of type (κ_v, λ_v). This holds for $y = 1, 2, \cdots, r$, and it follows that the principal right ideal of $\bar{\mathfrak{a}}$ generated by

---

[20] [11], p. 657.

[21] [5], (5).

$\bar{\alpha} = \sum_x h^x_{cd}(\alpha)\beta_x \neq 0$ is mapped on 0 by $\bar{\varphi}$, which gives a contradiction. Then $\mathfrak{a}$ is symmetric if $\bar{\mathfrak{a}}$ is.

If $\mathfrak{a}$ is symmetric, let φ be a symmetric function of $\mathfrak{a}$ which does not map any right ideal of $\mathfrak{a}$ on 0. If the corresponding function $\bar{\varphi}$ of $\bar{\mathfrak{a}}$ maps the principal right ideal of $\mathfrak{a}$ generated by

$$\bar{\alpha} = \sum_u c_u\beta_u \neq 0$$

on 0, then the principal right ideal of $\mathfrak{a}$ generated by $\bar{\alpha}$ would be mapped on 0 by the symmetric function φ. This shows that $\bar{\mathfrak{a}}$ is also symmetric.

III. Regular Representations

11. In this section a study is made of the simple parts of the regular representations $\mathfrak{R}$ and $\mathfrak{S}$ of $\mathfrak{a}$. By the method used in [11] it is possible to obtain a reduced form[22] of the regular representations by using as a basis the Cartan basis system ordered in a suitable manner. For the sake of brevity, the computations will be made only for the basic algebra $\bar{\mathfrak{a}}$; from the results for $\bar{\mathfrak{a}}$ it is easy to infer what the corresponding computations for $\mathfrak{a}$ itself would give.

To obtain a reduced form for the second regular representation $\bar{\mathfrak{S}}$ of $\bar{\mathfrak{a}}$, the basis elements β_u, $u = 1, 2, \cdots, r$ are arranged as follows:

(i) The β_u are classified according to the blocks to which they belong.[23]

(ii) Sets B_ρ are formed by taking all β_u such that $\lambda_u = \rho$.

(iii) In the set B_ρ are taken first the elements which do not belong to the radical (there is only one, namely $e_{\rho,11}$), then those which belong to $\bar{\mathfrak{n}}$, followed by those which belong to $\bar{\mathfrak{n}}^2$, and so on.

Let us use p_ρ as subscript for the first β_u of B_ρ, and q_ρ for the last element, so that $\beta_{p_\rho} = e_{\rho,11}$.

A detailed discussion of the splitting, under such an arrangement of the basis elements, of the regular representation into indecomposable and irreducible constituents has been given in [11] so here only the results will be stated. As a consequence of (i), the second regular representation, $\bar{\mathfrak{S}}$, of $\bar{\mathfrak{a}}$ decomposes into parts $\bar{\mathfrak{S}}_i$ which in fact are the regular representations of indecomposable two-sided ideals. Each set B_ρ forms a basis for a right ideal of $\bar{\mathfrak{a}}$, and the $\bar{\mathfrak{S}}_i$ decompose into indecomposable constituents $\bar{\mathfrak{V}}_\rho$ corresponding to these sets B_ρ. By the arrangement (iii) each such indecomposable constituent $\bar{\mathfrak{V}}_\rho$ of $\bar{\mathfrak{S}}$ is split according to its upper Loewy constituents[24] and is in reduced form.

The elementary modules $\bar{\mathfrak{H}}_u$ of $\bar{\mathfrak{a}}$ are of first degree, that is, they consist of 1×1 matrices. The elementary modules $\bar{\mathfrak{H}}_{p_\rho}$ corresponding to the $\beta_{p_\rho} = e_{\rho,11}(\rho = 1, 2, \cdots, k)$ are the irreducible representations of $\bar{\mathfrak{a}}$. We shall now study the matrix in $\bar{\mathfrak{V}}_\rho$ corresponding to the element $\bar{\alpha} = \sum_{x=1}^r \bar{H}_x(\bar{\alpha})\beta_x$. Let

---

[22] For the statement of the meaning of reduced form of a representation, see [12], Introduction.

[23] β_u, β_v belong to the same block if there is a chain $\beta_u, \beta_a, \cdots, \beta_b, \beta_v$, such that neighboring elements have at least one type index in common.

[24] [11], Th. 3.

$\beta_v \in B_\rho$, then

$$
(31)\qquad
\begin{aligned}
\bar{\alpha}\beta_v &= \sum_{y=p_{\kappa_v}}^{q_{\kappa_v}} \bar{H}_y(\bar{\alpha})\beta_y\beta_v = \sum_{u=p_\rho}^{q_\rho}\left(\sum_{y=p_{\kappa_v}}^{q_{\kappa_v}} \bar{H}_y(\bar{\alpha})c_{yvu}\right)\beta_u \\
&= \sum_u \bar{E}_{uv}(\bar{\alpha})\beta_u .
\end{aligned}
$$

We observe that $c_{yvu} = 0$ except when $(\kappa_y, \lambda_y) = (\kappa_u, \kappa_v)$. Further, if we denote by τ_x the power of the radical to which β_x belongs, then from (31), $\tau_y + \tau_v \leqq \tau_u$. Then $c_{yvu} = 0$ unless $\tau_y \leqq \tau_u - \tau_v$. This gives that $\bar{E}_{uv}(\bar{\alpha}) = \sum_w c_{wvu}\bar{H}_w(\bar{\alpha})$ where $(\kappa_w, \lambda_w) = (\kappa_u, \kappa_v)$ and $\tau_w \leqq \tau_u - \tau_v$. In particular, when $u = v$, $(\kappa_w, \lambda_w) = (\kappa_v, \kappa_v)$ and $\tau_w = 0$, hence $w = p_{\kappa_v}$, and $\bar{E}_{vv}(\bar{\alpha}) = \bar{H}_{p_{\kappa_v}}(\bar{\alpha})$. Also, when $v = p_\rho$, (31) becomes

$$
\bar{\alpha}\cdot\beta_{p_\rho} = \bar{\alpha}e_{\rho,11} = \sum_{x=p_\rho}^{q_\rho} \bar{H}_x(\bar{\alpha})\beta_x
$$

so that $\bar{E}_{up_\rho}(\bar{\alpha}) = \bar{H}_u(\bar{\alpha})$.

The module $\bar{\mathfrak{E}}_{uv}$, consisting of the matrices $\bar{E}_{uv}(\bar{\alpha})$, is a simple part[25] of the indecomposable constituent $\bar{\mathfrak{V}}_\rho$ of $\bar{\mathfrak{S}}$ corresponding to the basis elements of B_ρ. The computation of the regular representation of $\mathfrak{a}$ merely replaces the elementary modules $\bar{\mathfrak{H}}_w$ of the basic algebra $\bar{\mathfrak{a}}$ by the corresponding elementary modules $\mathfrak{H}_w$ of $\mathfrak{a}$, and so we obtain

11.1. *The indecomposable constituent $\mathfrak{V}_\rho$ of the regular representation $\mathfrak{S}$ of $\mathfrak{a}$ may be written with simple parts $\mathfrak{E}_{uv}$, $(u = p_\rho, p_\rho + 1, \cdots, q_\rho;\, v \leqq u)$,*

$$
\mathfrak{V}_\rho = \begin{Bmatrix}
\mathfrak{E}_{p_\rho p_\rho} & & & \\
\cdot & \cdot & & \\
\cdot & & \cdot & \\
\cdot & & & \cdot \\
\mathfrak{E}_{q_\rho p_\rho} & \cdot & \cdot\ \cdot & \mathfrak{E}_{q_\rho q_\rho}
\end{Bmatrix}
$$

where the simple part $\mathfrak{E}_{uv}$ is of type (κ_u, κ_v), and is a linear combination $\sum_w c_{wvu}\mathfrak{H}_w$ of elementary modules $\mathfrak{H}_w$ of type (κ_u, κ_v) which belong to $\mathfrak{n}^{\tau_w}$, $\tau_w \leqq \tau_u - \tau_v$. In particular, $\mathfrak{E}_{vv} = \mathfrak{H}_{p_{\kappa_v}}$, $\mathfrak{E}_{up_\rho} = \mathfrak{H}_u$.

Remarks: Let $t = \tau_{q_\rho}$, that is, t is the highest power of the radical which contains a basis element of the set B_ρ. Let $\mathfrak{L}_h$ denote the upper Loewy constituent of $\mathfrak{V}_\rho$ which corresponds to $\mathfrak{n}^h$. Then if we split $\mathfrak{V}_\rho$ according to its upper Loewy constituents, we obtain

$$
\mathfrak{V}_\rho = \begin{Bmatrix}
\mathfrak{L}_0 & & & & \\
\mathfrak{L}_{1,0} & \mathfrak{L}_1 & & & \\
\cdot & & \cdot & & \\
\cdot & & & \cdot & \\
\cdot & & & & \cdot \\
\mathfrak{L}_{t,0} & \mathfrak{L}_{t,1} & \cdot & \cdot\ \cdot & \mathfrak{L}_t
\end{Bmatrix}
$$

[25] For the definition of simple part of a matrix algebra, see [12], Introduction.

where the non-zero simple parts of $\mathfrak{L}_{m,j}$ belong to $\mathfrak{n}^i$ with $i \leqq m - j$. The first Loewy constitueht $\mathfrak{L}_0$ is the irreducible representation $\mathfrak{H}_{p_\rho}$ of $\mathfrak{a}$[26] and from 11.1 it follows that $\mathfrak{L}_{h,0}$ is of form

$$\mathfrak{L}_{h,0} = \begin{pmatrix} \mathfrak{H}_{u_1} \\ \cdot \\ \cdot \\ \cdot \\ \mathfrak{H}_{u_r} \end{pmatrix}$$

where $\beta_{u_1}, \cdots, \beta_{u_r}$ are the elements of B_ρ which belong to $\bar{\mathfrak{n}}^h$. Then there exist elements of $\mathfrak{a}$ for which the corresponding matrices in $\mathfrak{B}_\rho$ are of form

$$(32) \qquad \begin{pmatrix} \mathfrak{L}_{t-k,0} & & & & & \\ \cdot & \cdot & & & & \\ \cdot & & \cdot & & & \\ \cdot & & & \cdot & & \\ \mathfrak{L}_{t-1,0} & \cdot & \cdot & \cdot & \mathfrak{L}_{t-1,k-1} & \\ \mathfrak{L}_{t,0} & \cdot & \cdot & \cdot & \cdot & \mathfrak{L}_{t,k} \end{pmatrix}$$

where the parts $\mathfrak{L}_{t-\sigma,0}$ in the first column may be arbitrarily chosen. If $\mathfrak{a}$ is a quasi-Frobeniusean algebra[27] then $\mathfrak{L}_{t,0}$ contains just one elementary module of type (ρ, ρ^*) (where * denotes a permutation of $1, 2, \cdots, k$). If $\mathfrak{a}$ is symmetric, then $\mathfrak{L}_{t,0}$ consists of one elementary module of type (ρ, ρ).

Analogous statements may be made in regard to the simple parts of the indecomposable constituents of the first regular representation.

THE UNIVERSITY OF MICHIGAN.

BIBLIOGRAPHY

1. A. A. ALBERT, *Structure of Algebras*, American Mathematical Society Colloquium Publications, vol. 24, 1939.
2. R. BRAUER, *Uber die Darstellung von Gruppen in Galoisschen Feldern*, Actualites scientifigues et industrielles, No. 195 (1935).
3. R. BRAUER, *On sets of matrices with coefficients in a division ring*, Transactions of the American Mathematical Society, vol. 49 (1941), pp. 502–548.
4. R. BRAUER, *On the modular representations of algebras*, Proceedings of the National Academy of Sciences, vol. 25, (1939), p. 290.
5. R. BRAUER, C. NESBITT, *On the regular representations of algebras*, Proceedings of the National Academy of Sciences, vol. 23 (1937), pp. 430–434.
6. R. BRAUER, C. NESBITT, *On the modular representations of groups of finite order*, University of Toronto Studies, Mathematical Series, vol. 4, 1937.
7. R. BRAUER, C. NESBITT, *On the modular characters of groups*, Annals of Mathematics, vol. 42 (1941), pp. 556–590.

---

[26] [11], Th. 2.

[27] For the definition of quasi-Frobeniusean algebra, see [9].

8. MARSHALL HALL, *The position of the radical in an algebra*, Transactions of the American Mathematical Society, vol. 48, (1940), pp. 391–404.
9. T. NAKAYAMA, *On Frobeniusean algebras*, I, Annals of Mathematics, vol. 40, (1939), pp. 611–633.
10. T. NAKAYAMA, C. NESBITT, *Note on symmetric algebras*, Annals of Mathematics, vol. 39 (1938), pp. 659–668.
11. C. NESBITT, *On the regular representations of algebras*, Annals of Mathematics, vol. 39 (1938), pp. 634–658.
12. W. M. SCOTT, *On matrix algebras over an algebraically closed field*, Annals of Mathematics, vol. 43 (1942) pp. 147–160.

ANNALS OF MATHEMATICS
Vol. 44, No. 3, July, 1943

OPEN ADDITIVE SEMI-GROUPS OF COMPLEX NUMBERS*

BY EINAR HILLE AND MAX ZORN

(Received March 3, 1943)

The main object of the present paper is to determine the open, connected, additive semi-groups of the complex number plane. These sets are important as parameter manifolds of one-parameter semi-groups of linear transformations which have been studied in detail by one of us.[1] It is shown that these parameter manifolds depend on the upper semi-continuous solutions of the inequality[2]

$$\varphi(\xi_1 + \xi_2) \leqq \varphi(\xi_1) + \varphi(\xi_2),$$

the function φ being defined on the ray $\xi > 0$ or on the whole line. The semi-group consists of all numbers $\xi + i\eta$ with $\eta > \varphi(\xi)$.

Finally we show that every open semi-group is the maximal domain of existence of a one-parameter analytic semi-group of linear transformations defined on a suitable Banach space.

It is obvious from the definition of a semi-group (see below) that the complex numbers may be replaced by any two-dimensional vector space over the real number field. Furthermore it turns out that our considerations apply with minor modifications to n-dimensional vector spaces, and many intermediate definitions and results are valid in still more general cases. We may present this more general form of the theory, together with a study of closed semi-groups and with additional facts about the structure of open plane semi-groups at another occasion.

1. Definition and structural properties

Dealing with complex numbers as elements of a two-dimensional vector space $A = A_2$ over the reals we define

1.1 *A semi-group S of the (group-) space A is a subset of A which* (i) *is additive,* (ii) *possesses in every neighborhood of* 0 *an element different from* 0.[3,4]

---

* Presented to the American Mathematical Society November 28, 1942.

[1] See E. Hille [5] in the list of references at the end of the present paper.

[2] This inequality has been studied by R. Cooper [4]. He assumes $\varphi(\xi)$ defined for all ξ and proves that a solution which is monotonic for large positive (negative) ξ has the property that $\varphi(\xi)/\xi$ tends to a finite limit when $\xi \to +\infty(-\infty)$. He also shows that an odd solution is of the form $k\xi$ and an even solution is nowhere negative and gives examples of discontinuous measurable solutions.

[3] There exist additive sets which do not satisfy (ii). Assuming the set to be a convex domain, E. Hille has shown that it is the maximal domain of existence of a one-parameter analytic semi-group of linear transformations on $L_1(-\pi, \pi)$ to itself. His construction is entirely different from that given in §3 below.

[4] If the group space is finite dimensional, $A = A_n$, we suppose that a metric is introduced in A_n, for instance the ordinary Euclidean one, and that the metric defines the neighborhood topology. We denote the length of a by $|a|$. In the present paper $n = 1$ or 2.

The closure $\bar{S}$ of a semi-group S is likewise a semi-group. The following observation—which is capable of considerable generalization—is basic for our theory.

1.2 *To every semi-group S of A there exists a vector $b \neq 0$ such that the ray ρb, $\rho \geqq 0$, is in the closure of S.*

Indeed, there must exist a sequence $\{a_j\}$, $a_j \neq 0$, $a_j \in S$ with $\lim a_j = 0$. The unit vectors $a_j/|a_j|$ must have at least one limit point in A and without loss of generality we may assume $\lim a_j/|a_j| = b$ where $1 = |b| \neq 0$. For $\rho \geqq 0$ we determine a sequence of integers n_j such that $n_j > 0$, $\lim n_j |a_j| = \rho$, for instance, $n_j = [\rho/|a_j|] + 1$. The relation $\rho b = (\lim n_j |a_j|)(\lim a_j/|a_j|) = \lim n_j a_j$ shows that ρb is limit of a sequence from S.

1.3 *An open semi-group S is the interior of its closure $\bar{S}$.*

In other terms: if a point of A has a neighborhood on which S is dense it must be an element of S. Or, if a point x of A does not belong to S, then every neighborhood U_x contains an open set which is not void and has no points in common with S.

To prove this we take a neighborhood U_0 of the 0-element such that $x - U_0$ is in U_x; in this neighborhood there will be a vector y which, together with a whole neighborhood U_y, is contained in both S and U_0. The non-void open set $x - U_y$ is contained in U_x, but has no points in S; for if $x - u$ were in S, $x = u + (x - u)$ would be, which is not true. This proves our lemma.

We note that the proof of this lemma extends literally to Hausdorff groups, but in this paper we shall use it only for A_2 and A_1.

An important consequence is

1.4 *For an open semi-group S we have $S + \bar{S} \subset S$.*

Since the closure of a semi-group is a semi-group, $S + \bar{S}$ is first of all contained in $\bar{S}$; and with every vector $s + \bar{s}$ we find a whole neighborhood $U_s + \bar{s}$ in this set; in other words, every point of $S + \bar{S}$ is an interior point of $\bar{S}$ and therefore contained in S itself.

1.5 *For an open semi-group S at least one vector $b \neq 0$ exists such that with s all vectors $s + \rho b$, $\rho \geqq 0$, are elements of S.*

It suffices to choose the vector b in $\bar{S}$ which is furnished by 1.2 and then to apply 1.4. Since S is open, a whole neighborhood of s and in particular all points $s + \eta' b$ with $|\eta'| < \epsilon(s)$, for suitable $\epsilon(s) > 0$, will belong to S. We may therefore state:

1.5.1 *For an open semi-group S there exists a vector $b \neq 0$ and a function $\epsilon(s) > 0$ such that with s all vectors $s + \eta' b$, $\eta' > -\epsilon(s)$, are contained in S.*

2. Representation by inequalities

For the sequel it will be useful to introduce the symbol $-\infty$ with the conventions $-\infty + (-\infty) = -\infty + \eta = -\infty$ and $\eta > -\infty$ for all real numbers η. Using φ as a common letter for $-\infty$ and reals we state without proof:

2.1.1 *A set $\{\eta\}$ of real numbers which contains with every η all η' with $\eta' > \eta - \epsilon(\eta)$, $\epsilon(\eta) > 0$, may always be described by an inequality $\eta > \varphi$.*

2.1.2 *An inequality* $\varphi_1 + \varphi_2 \geqq \varphi_3$ *is equivalent with the fact that* $\eta_1 > \varphi_1$, $\eta_2 > \varphi_2$ *implies* $\eta_1 + \eta_2 > \varphi_3$.

Returning to the open semi-groups we introduce a vector a, not colinear with b, such that every vector may be written in the form $\xi a + \eta b$.[5] For every element $s = \xi a + \eta b$ of S all vectors $s' = \xi a + \eta' b$, for a suitable $\epsilon(s) > 0$ and all $\eta' > \eta - \epsilon(s)$ are within S. The ordinates η of the elements in S which have the abscissa ξ fulfill thus the condition of 2.1.1, and we may say:

2.2 *An open semi-group consists of all vectors* $\xi a + \eta b$, *where* ξ *runs through a subset* T *of* A_1 *and* η *satisfies* $\eta > \varphi(\xi)$.

We denote a set which in this manner is determined by a set T and a function φ on T as the "restricted product" (T, φ).

2.3 *In order that the set* (T, φ) *be additive it is necessary and sufficient that*

(i) T *is additive,*

(ii) $\varphi(\xi)$ *satisfies the functional inequality*

2.3.1
$$\varphi(\xi_1 + \xi_2) \leqq \varphi(\xi_1) + \varphi(\xi_2).$$

If $\xi_i a + \eta_i b$ are in (T, φ) then $(\xi_1 + \xi_2)a + (\eta_1 + \eta_2)b$ is in (T, φ) exactly if (i) $\xi_1 + \xi_2$ is in T, (ii) $\eta_1 + \eta_2 > \varphi(\xi_1 + \xi_2)$ is implied by $\eta_i > \varphi(\xi_i)$. The first of these conditions is additivity of T, the second is equivalent with the inequality 2.3.1 by virtue of 2.1.2.

2.4 *In order that the restricted product* (T, φ) *be open in* A *it is necessary and sufficient that*

(i) T *is an open set in* $A_1 = \{\xi\}$,

(ii) *the function* $\varphi(\xi)$ *is upper semi-continuous in the sense that for real* σ *the set of all* ξ *for which* $\varphi(\xi) < \sigma$ *be open in* A_1.

Note that by virtue of (i) we might ask instead that the set be open in T. If now (T, φ) is open, and $\eta > \varphi(\xi)$ there will be numbers ϵ, $\delta > 0$ such that for $|\xi' - \xi| < \epsilon$, $|\eta' - \eta| < \delta$ the vectors $\xi' a + \eta' b$ are in (T, φ), or $\eta' > \varphi(\xi')$. The numbers ξ' are in T and constitute an A_1-neighborhood of ξ, which shows that T is open in A_1. If η is less than a real number σ we see that in this neighborhood $\varphi(\xi')$ is less than σ by choosing $\eta' = \eta$ in $\eta' > \varphi(\xi')$. Conversely, if conditions (i) and (ii) are satisfied, and for $\eta > \varphi(\xi)$, let us choose a $\delta > 0$ such that $\varphi(\xi) < \eta - \delta$; since φ is supposed to be upper semi-continuous and T open there will be a whole neighborhood $|\xi' - \xi| < \epsilon$ which is in T and where $\varphi(\xi') < \eta - \delta$. Therefore $\varphi(\xi') < \eta'$ will be true for $|\xi' - \xi| < \epsilon$, $|\eta' - \eta| < \delta$, which means that (T, φ) is open.

2.5 *In order that* (T, φ) *have in every neighborhood of* 0 *an element* $\neq 0$ *it is necessary and sufficient that*

(i) *there exists at least one point of* T *in every* A_1*-neighborhood of* 0,

(ii)
$$\varlimsup_{\xi \to 0} \varphi(\xi) \leqq 0.$$

---

[5] If we work with complex numbers we may apply a rotation and if necessary a reflection to S such that $a = 1$, $b = i$, and such that there are points in S with abscissas greater than zero.

We are using the definition $\underline{\lim}_{\xi\to 0}\, \varphi(\xi) = \sup_{\epsilon>0} \inf_{|\xi|<\epsilon} \varphi(\xi)$ and leave the proof of this lemma to the reader. Note that the first condition does not necessarily produce an infinity of points in every A_1-neighborhood of 0; this will however be the case if the set T is open in A_1. Altogether the conditions marked (i), valid jointly if (T, φ) is or is to be an open semi-group, add up to the statement that T is an open semi-group of the one dimensional vector space A_1.

2.6 *An open semi-group T of A_1 is one of the three sets: $\xi > 0$, $\xi < 0$, $\xi \epsilon A_1$.*

PROOF. The closure $\overline{T}$ of T contains a ray $\rho(\pm 1)$, $\rho \geqq 0$, according to 1.2; if it contains an additional element $\nu(\pm 1)$, $\nu < 0$, it contains *all* real numbers $\xi = (n\nu + \rho)(\pm 1)$. (Note that we have just determined all closed semi-groups of A_1.) In view of the generalization of 1.3 the semi-group will be the interior, with respect to A_1, of one of these three sets, which proves our theorem.

Combining this with the conditions marked (ii) we obtain the following description of the open semi-groups.

2.7 *After introduction of a suitable basis a, b an open semi-group consists of all vectors $\xi a + \eta b$ with $\eta > \varphi(\xi)$ where*

(i) *ξ varies over one of the three sets $\xi > 0$, $\xi < 0$, or A_1,*

(ii) *the function φ has real numbers or $-\infty$ as values, is upper semicontinuous, satisfies the inequality* 2.3.1 *and the condition $\underline{\lim}_{\xi\to 0}\, \varphi(\xi) \leqq 0$. Vice versa, any restricted product satisfying conditions* (i) *and* (ii) *will constitute an open semi-group of A_2.*

In this paper we record only such information about semi-groups as we need for the analytic developments of the following section. The next four statements suffice for our purposes.

2.8 *An open semi-group S is connected.*

We show right away that S is arc-wise connected by exhibiting for any two elements $\xi_i a + \eta_i b$ of S a polygon which connects them and is contained in S. The construction is possible because the upper semicontinuous function has an upper bound $\mu > \varphi(\xi)$ on the interval $\xi_1 \leqq \xi \leqq \xi_2$. The polygon is then made up by the following three—possibly degenerate—arcs:

$$\begin{aligned} &\text{(i)} \quad \xi = \xi_1,\ \eta = \eta_1 + \theta(\mu - \eta_1); \\ &\text{(ii)} \quad \xi = \xi_1 + \theta(\xi_2 - \xi_1),\ \eta = \mu; \qquad\qquad 0 \leqq \theta \leqq 1. \\ &\text{(iii)} \quad \xi = \xi_2,\ \eta = \mu + \theta(\eta_2 - \mu). \end{aligned}$$

2.9 *An open semi-group S is simply connected.*

Let C be a simple closed curve consisting of points in S; we have to show that a point c inside of C, that is, a point which cannot be connected with arbitrarily far points by an arc which does not intersect C, is necessarily contained in S. Indeed, consider the ray $c - \rho b$, $\rho \geqq 0$. This ray will intersect C at a point $s_0 = c - \rho_0 b$ of S, and that implies, by virtue of 1.5, that $c = s_0 + \rho_0 b$ is in S.

2.10 *If S is not the whole plane A_2 there exist points whose distance from the set S is > 0.*

That is so because S is the interior of its closure.

2.11 *If S is not the whole plane it does not contain the* 0*-element.*
For with 0 all sufficiently small vectors would have to be in S, and every vector is a multiple of arbitrarily small vectors.

3. Maximal parameter sets

In this paragraph we shall prove an existence theorem for analytical semi-groups of linear bounded transformations on a Banach space to itself. Some preliminary explanations and definitions of the concepts involved are in order.

Let E be a Banach space of elements $x, y, \cdots$, and let E^* be the adjoint space of linear bounded functionals defined on E. Let S be a set of complex numbers $s = \sigma + i\tau$ and let $\{T_s\}$ be a one-parameter family of linear bounded transformations on E to E defined for $s \in S$.

3.1. *$\{T_s\}$ is said to be a semi-group if*

(i) *the set S is additive, and*

(ii) $T_sT_tx = T_tT_sx = T_{s+t}x$ *for all $x \in E$ and all $s, t \in S$.*

3.2. *T_s is said to be holomorphic in S if* (i) *S is a domain, and* (ii) *the complex valued functions $L(T_sx)$ are holomorphic in S for all $x \in E$ and all $L \in E^*$. S is said to be the maximal domain of analytic existence of T_s if every accessible boundary point of S is a singular point of at least one of the functions $L(T_sx)$.*[6]

This definition of maximal domain of existence disregards completely the possibility of quasi-analytic or more general non-analytic continuation of T_s valid for all elements of E. We also disregard the possibility of finding an analytic continuation of T_s valid on some subspace of E. Simple examples of both possibilities can be found.

N. Dunford[7] has found a simple criterion for holomorphism in S:

3.3. *A necessary and sufficient condition in order that T_s be holomorphic in S is that the difference quotient $(1/h)(T_{s+h} - T_s)x$, $s \in S$, $s + h \in S$, shall converge strongly to a limit when $h \to 0$* for *every $x \in E$.*

We shall now prove[8]

3.4. *Given an open additive set S in the complex plane such that $s = 0$ belongs to $\bar{S} - S$, i.e., S is an open semi-group in the sense of definition* 1.1. *Then there exists a Banach space E and a one-parameter family $\{T_s\}$ of bounded linear transformations on E to E defined for s in S such that* (i) *$\{T_s\}$ is a semi-group in the sense of* 3.1, (ii) *T_s is holomorphic in S, and* (iii) *S is the maximal domain of analytic existence of T_s.*

That S is a domain was stated in 2.8.

We shall give two slightly different constructions. In the first we take for E the set of all functions $f(z)$ bounded and holomorphic in S and define $||f(z)|| =$

---

[6] That is, to the accessible boundary point s_0 should correspond at least one function $L(T_sx)$ and at least one rectifiable arc C in S ending at s_0 such that the radius of convergence of the Taylor expansion of $L(T_sx)$ about $s = s_1$ on C tends to zero when $s_1 \to s_0$ along C.

[7] See E. Hille [5], pp. 6–7.

[8] We exclude the case in which S is the whole finite plane since the existence of analytical groups is well known.

$\sup_{z \in S} | f(z) |$. This is a normed linear vector space complete in its metric. We then define $T_s f(z) = f(z + s)$ for $s \in S$. This is clearly a semi-group in the sense of 3.1. In order to prove (ii) we use the criterion given in 3.3. We have the lemma

3.5. *If s and $s + h$ belong to S, we have*

$$|| (1/h)[f(z + s + h) - f(z + s)] - f'(z + s) || \leq 2 | h | [\delta(s)]^{-2} || f(z) ||$$

for $| h | < \frac{1}{2}\delta(s)$ where $\delta(s)$ is the greatest lower bound of the distance of $z + s$ from the boundary of S when z ranges over S and s is fixed in S.

The lemma can be proved with the aid of Schwarz's lemma but Cauchy's integral gives a more direct proof. We have

$$\frac{1}{h} [f(z + s + h) - f(z + s)] - f'(z + s) = \frac{1}{2\pi i} \int_C \frac{f(w)\, dw}{(w - z - s)^2(w - z - s - h)}$$

where C is a circle interior to S with center at $w = z + s$ which also contains $z + s + h$. From this representation the inequality is an immediate consequence and 3.5 implies 3.4 (ii).

We note that S is a simply-connected domain which omits the interior and boundary of a suitably chosen circle, by 2.10. Hence E contains all functions which are bounded and holomorphic outside a suitably chosen circle. Of greater interest to us is the fact that E contains elements which have S as their natural domain of existence. Such elements can be constructed as follows. Let $w = F(z)$ be a function which maps S conformally on the interior of the unit-circle, $| w | < 1$, and let $\sum_0^\infty a_n w^n$ be a power series having $| w | = 1$ as its natural boundary and such that $\sum_0^\infty | a_n | < \infty$. Then $\sum_0^\infty a_n [F(z)]^n$ is an element of E having S as its natural domain of existence.

In order to prove (iii) we now argue as follows. We denote by E_0 the linear sub-space consisting of all functions $f(z)$ in E such that $\lim_{n \to \infty} f(z_n)$ exists where $z_n \in S$ and $z_n \to 0$ when $n \to \infty$, $\{z_n\}$ being a fixed preassigned but otherwise arbitrary point-set. We denote this limit arbitrarily by $f(0)$ when it exists. Then the convention $L_0[f(z)] = f(0)$ defines a linear bounded functional on E_0 such that $| L_0[f] | \leqq || f ||$. We note that for a fixed s in S and any $f(z)$ in E we have $f(z + s) \in E_0$ so that $L_0[f(z + s)] = f(s)$. The Hahn-Banach theorem having been generalized to complex-valued functionals by F. Bohnenblust and A. Sobczyk [3], we can extend the functional L_0 with unchanged norm to all elements of E. We have consequently $L_0[T_s f(z)] = f(s)$ for every $f(z) \in E$ and all s in S. It is obvious that $L_0[T_s f(z)]$ is holomorphic in S as is required by 3.2. But we have just seen that there are elements of E having S as their natural domain of existence. For such a choice of $f(z)$ the function $L_0[T_s f(z)] = f(s)$ has S as its natural domain of existence. It follows that S is the maximal domain of analytic existence of T_s. This completes the argument in the first case.

For the second construction we consider instead the space E_1 of functions $f(z)$ holomorphic in S such that

$$|| f(z) || = \left\{ \int\int_S | f(z) |^2 \, d\omega \right\}^{\frac{1}{2}} < \infty.$$

We note that if $\varphi(z) \in E$, the space of bounded holomorphic functions in S, and if $z = a$ has positive distance from S, then $\varphi(z)\,(z - a)^{-2} \in E_1$. Furthermore, this function will have S as its natural domain of existence if $\varphi(z)$ has this property. It is easy to see that E_1 is a Hilbert space and a fortiori a complex Banach space.

From this it follows that every bounded linear functional defined on E_1 is given by the formula

$$L[f(z)] = \iint_S f(z)\overline{g(z)}\, d\omega,$$

where $g(z)$ is any element of E_1. A particular case of this formula should be noted. It has been shown by S. Bergmann [1] and S. Bochner [2] that there exists a kernel $K(t, z)$ such that

$$f(z) = \iint_S K(t, z) f(t)\, d\omega,$$

where the integration is taken with respect to t. Here $f(z)$ is any element of E_1, $K(t, z)$ depends only upon S but not upon $f(z)$, $\overline{K(t, z)} = K(z, t)$, and for fixed t in S, $K(t, z)$ is an element of E_1. Hence we may choose $g(z) = K(\alpha, z)$, where α is a fixed point in S, and obtain a functional L_α such that $L_\alpha[f(z)] = f(\alpha)$.

We now define $T_s f(z) = f(z + s)$ for $f(z) \in E_1$, $s \in S$. This definition satisfies 3.4 (i). But

$$L[T_s f(z)] = \iint_S f(z + s)\overline{g(z)}\, d\omega$$

and this is clearly a holomorphic function of s in S so that (ii) also holds.

Suppose now that s_0 is an accessible boundary point of S and suppose that the maximal domain of analytic existence of T_s should include s_0.[9] Then there must exist a small fixed neighborhood of $s = s_0$ in which all the functions $L[T_s f(z)]$ are holomorphic. But if we choose in particular $L = L_\alpha$, then $L_\alpha[T_s f(z)] = f(\alpha + s)$, and all these functions must be holomorphic in the neighborhood in question, no matter how we choose α in S and $f(z)$ in E_1. But this is impossible as we see by taking for $f(z)$ a function having S as its natural domain of existence and choosing α sufficiently near to the origin. Hence S is the maximal domain of analytical existence of T_s and the second construction has been completely verified.

References

[1] S. Bergmann, *Über die Entwicklung der harmonischen Funktionen der Ebene und des Raumes nach Orthogonalfunktionen*, Mathematische Annalen, 86 (1922) 238–271.

[2] S. Bochner, *Über orthogonale Systeme analytischer Funktionen*, Mathematische Zeitschrift, 14 (1922) 180–207.

[9] If S has been normalized through rotation in such a manner that $a = 1$, $b = i$ in the representation 2.7, then every vertical line $x = \xi$ contains an accessible boundary point of S.

[3] H. F. Bohnenblust and A. Sobczyk, *Extensions of Functionals on Complex Linear Spaces*, Bulletin of the American Mathematical Society, 44 (1938), 91–93.
[4] R. Cooper, *The Converses of the Cauchy-Hölder Inequality and the Solutions of the Inequality* $g(x + y) \leq g(x) + g(y)$. Proceedings of the London Mathematical Society, (2) 26 (1926–27) 415–432.
[5] E. Hille, *Notes on Linear Transformations.* II. *Analyticity of Semi-Groups.* Annals of Mathematics, (2) 40 (1939) 1–47.

Yale University,
University of California.

Annals of Mathematics
Vol. 44, No. 3, July, 1943

ON A PROJECTIVE INVARIANT OF A NON-HOLONOMIC SURFACE

By Hsien-Chung Wang*

(Received November 9, 1942)

1. Introduction. The purpose of this paper is to give a projective invariant of a non-holonomic surface in a three-dimensional space. This invariant can be regarded as an analogue of the projective linear element of an ordinary surface, since it has a geometrical meaning similar to the latter. Analytically, as we shall show below, it is the quotient of the product of two quadratic differential forms by the square of a Pfaffian form. While there are projectively distinct ordinary surfaces which have the same projective linear element,[1] we shall prove in our case the main theorem that two non-holonomic surfaces have the same invariant when and only when there exists a collineation or a correlation carrying one surface to the other.

2. Canonical frames and the projective linear element of a non-holonomic surface. Let x, y, z be a set of coordinates in a projective space of three dimensions. A non-holonomic surface S is defined by an equation of the form[2]

$$A(x, y, z)dx + B(x, y, z)dy + C(x, y, z)dz = 0. \tag{1}$$

Geometrically it associates to every point A of the space a plane π through the point. The plane π is called the *tangent plane* of S at A. There are in general two directions on π through A having the property that the tangent plane at a neighboring point A' of A on any of these two directions intersects π along AA'. We assume these two directions to be real and distinct and call them the *asymptotic directions*. The two lines through A along the asymptotic directions are called the *asymptotic tangents*.[3]

Following Cartan we take as a projective frame[4] a set of four linearly independent analytic points A, A_1, A_2, A_3 such that $|AA_1A_2A_3| = 1$. To each point of the space we attach the most general family of frames $AA_1A_2A_3$ satisfying the conditions: 1). A coincides with the point; 2). AA_1, AA_2 are the two asymptotic tangents. Such a frame depends on the coordinates of A and a set of "secondary parameters" which determine the frame within the sub-family attached to the same point A. The totality of these frames attached to different points of the space satisfies a system of equations of the form

---

* The author wishes to express his thanks to Professor Shiing-shen Chern for his valuable suggestions.

[1] E. Cartan, *Sur la déformation projective des surfaces*, Annales Ec. Norm. Sup., (3) 37(1920), pp. 259–356.

[2] E. Bompiani, *Sulle varietà anolonome*, 1, 2, Rend. R. Accad. Lincei, serie VI, 27(1938), pp. 37–52.

[3] E. Bompiani, ibid.

[4] E. Cartan, ibid.

$$
(2) \qquad \begin{cases} dA = \omega_0^0 A + \omega_0^1 A_1 + \omega_0^2 A_2 + \omega_0^3 A_3, \\ dA_1 = \omega_1^0 A + \omega_1^1 A_1 + \omega_1^2 A_2 + \omega_1^3 A_3, \\ dA_2 = \omega_2^0 A + \omega_2^1 A_1 + \omega_2^2 A_2 + \omega_2^3 A_3, \\ dA_3 = \omega_3^0 A + \omega_3^1 A_1 + \omega_3^2 A_2 + \omega_3^3 A_3 \end{cases}
$$

with the relation

$$
(3) \qquad \omega_0^0 + \omega_1^1 + \omega_2^2 + \omega_3^3 = 0.
$$

For simplicity we shall write, in what follows, ω^i for ω_0^i, $i = 1, 2, 3$. Since the lines AA_1, AA_2 remain fixed when A is fixed, it follows that ω_1^2, ω_2^1, ω_1^3, ω_2^3 are linear combinations of ω^1, ω^2, ω^3. But the asymptotic directions are given by

$$
\omega^3 = 0 \qquad \omega^1\omega_1^3 + \omega^2\omega_2^3 = 0
$$

so that ω_1^3 is a linear combination of ω^2, ω^3 and ω_2^3 a linear combination of ω^1, ω^3 only. We can therefore put

$$
(4) \qquad \begin{cases} \omega_1^3 = b\omega^2 + c\omega^3, & \omega_2^3 = b^*\omega^1 + c^*\omega^3, \\ \omega_1^2 = \alpha\omega^1 + \beta\omega^2 + \gamma\omega^3, & \omega_2^1 = \beta^*\omega^1 + \alpha^*\omega^2 + \gamma^*\omega^3. \end{cases}
$$

From this family of frames we shall choose a sub-family characterized by invariant conditions. Let us denote by δ the symbol of differentiation under which only the secondary parameters vary, the coordinates x, y, z of the point A remaining fixed, and by d a symbol under which all the variables vary. For simplicity of notation we write

$$
(5) \qquad \omega_i^j(D) = \omega_i^j \qquad \omega_i^j(\delta) = e_i^j \qquad i, j = 0, 1, 2, 3.
$$

Then we have

$$
(6) \qquad e_0^0 + e_1^1 + e_2^2 + e_3^3 = 0, \qquad e^1 = e^2 = e^3 = e_1^3 = e_2^3 = e_1^2 = e_2^1 = 0
$$

It is well-known that the Pfaffian forms in (2) satisfy the "equations of structure":

$$
(7) \qquad \begin{cases} (\omega^1)' = [(\omega_0^0 - \omega_1^1)\omega^1] + [\omega^2\omega_2^1] + [\omega^3\omega_3^1], \\ (\omega^2)' = [(\omega_0^0 - \omega_2^2)\omega^2] + [\omega^1\omega_1^2] + [\omega^3\omega_3^2], \\ (\omega^3)' = [(\omega_0^0 - \omega_3^3)\omega^3] + [\omega^1\omega_1^3] + [\omega^2\omega_2^3], \\ (\omega_1^3)' = [(\omega_1^1 - \omega_3^3)\omega_1^3] + [\omega_1^0\omega^3] + [\omega_1^2\omega_2^3], \\ (\omega_2^3)' = [(\omega_2^2 - \omega_3^3)\omega_2^3] + [\omega_2^0\omega^3] + [\omega_2^1\omega_1^3], \\ (\omega_1^2)' = [(\omega_1^1 - \omega_2^2)\omega_1^2] + [\omega_1^0\omega^2] + [\omega_1^3\omega_3^2], \\ (\omega_2^1)' = [(\omega_2^2 - \omega_1^1)\omega_2^1] + [\omega_2^0\omega^1] + [\omega_2^3\omega_3^1], \end{cases}
$$

from which it follows that

$$
(8)\qquad \begin{cases}
\delta\omega^1 = (e_0^0 - e_1^1)\omega^1 - e_3^1\omega^3, & \delta\omega^2 = (e_0^0 - e_2^2)\omega^2 - e_3^2\omega^3, \\
\delta\omega^3 = (e_0^0 - e_3^3)\omega^3, & \delta\omega_1^3 = (e_1^1 - e_3^3)\omega_1^3 + e_1^0\omega^3, \\
\delta\omega_2^3 = (e_2^2 - e_3^3)\omega_2^3 + e_2^0\omega^3, & \delta\omega_1^2 = (e_1^1 - e_2^2)\omega_1^2 + e_1^0\omega^2 - e_3^2\omega_1^3, \\
\delta\omega_2^1 = (e_2^2 - e_1^1)\omega_2^1 + e_2^0\omega^1 - e_3^1\omega_2^3.
\end{cases}
$$

Equations (4) and (8) then give

$$
(9)\qquad \begin{cases}
\delta b = (e_1^1 + e_2^2 - e_0^0 - e_3^3)b, & \delta c = e_1^0 + e_3^2 b + (e_1^1 - e_0^0)c, \\
\delta b^* = (e_1^1 + e_2^2 - e_0^0 - e_3^3)b^*, & \delta c^* = e_2^0 + e_3^1 b^* + (e_2^2 - e_0^0)c^*, \\
\delta\alpha = (2e_1^1 - e_0^0 - e_2^2)\alpha, & \delta\gamma = (e_1^1 + e_3^3 - e_0^0 - e_2^2)\gamma \\
& \qquad + e_3^1\alpha + e_3^2\beta - e_3^2 c, \\
\delta\alpha^* = (2e_2^2 - e_0^0 - e_1^1)\alpha^*, & \delta\gamma^* = (e_2^2 + e_3^3 - e_0^0 - e_1^1)\gamma^* \\
& \qquad + e_3^2\alpha^* + e_3^1\beta^* - e_3^1 c^*, \\
\delta\beta = (e_1^1 - e_0^0)\beta + e_1^0 - e_3^2 b, & \delta\beta^* = (e_2^2 - e_0^0)\beta^* + e_2^0 - e_3^1 b^*.
\end{cases}
$$

We shall suppose that the tangent plane of the non-holonomic surface depends on three essential parameters. Then we have

$$bb^* \neq 0.$$

Equations (9) show that we can choose the secondary parameters such that the following conditions are satisfied:

$$bb^* = 1, \qquad c = c^* = \beta = \beta^* = 0,$$

i.e.

$$
(10)\qquad \begin{cases}
\omega_1^3 = b\omega^2, & \omega_2^3 = \dfrac{1}{b}\omega^1, \\
\omega_1^2 = \alpha\omega^1 + \gamma\omega^3, & \omega_2^1 = \alpha^*\omega^2 + \gamma^*\omega^3.
\end{cases}
$$

To keep these conditions unaltered, we must have

$$
(11)\qquad e_1^1 + e_2^2 = e_0^0 + e_3^3 = e_3^1 = e_3^2 = e_1^0 = e_2^0 = 0.
$$

Equations (8) then become

$$
(12)\qquad \begin{cases}
\delta\omega^1 = (e_0^0 - e_1^1)\omega^1, & \delta\omega^2 = (e_0^0 - e_2^2)\omega^2, & \delta\omega^3 = 2e_0^0\omega^3, \\
\delta\omega_1^3 = (e_1^1 - e_3^3)\omega_1^3, & \delta\omega_2^3 = (e_2^2 - e_3^3)\omega_2^3, \\
\delta\omega_1^2 = 2e_1^1\omega_1^2, & \delta\omega_2^1 = -2e_1^1\omega_2^1.
\end{cases}
$$

The family of frames for which the conditions (10) are fulfilled is called the family of *canonical frames*. The canonical frames at a point A depend on three

secondary parameters. By making use of relations (12), we can verify that the expression

$$(13)\qquad d\sigma^2 = \frac{(\omega^2\omega_1^3 - \omega^3\omega_1^2)(\omega^1\omega_2^3 - \omega^3\omega_2^1)}{(\omega^3)^2}$$

is the quotient of two differential forms in x, y, z and is independent of the secondary parameters. It is therefore a projective invariant of the non-holonomic surface, which we shall call its *projective linear element.*

3. Geometrical interpretation of the projective linear element. To interpret the projective linear element geometrically, we employ a notation due to Fubini.[5] Let P_1, P_2 be two analytic points. We denote by $[P_1P_2]$ the analytic line with Plückerian coordinates equal respectively to the minors of the matrix $\| P_1P_2 \|$. Furthermore, if $[P_1P_2]$ and $[P_1'P_2']$ define two lines, then the scalar product $S[P_1P_2][P_1'P_2']$ stands for the determinant $| P_1P_2P_1'P_2' |$. Thus two lines $[P_1P_2]$, $[P_1'P_2']$ intersect if and only if

$$S[P_1P_2][P_1'P_2'] = 0.$$

For a canonical frame $AA_1A_2A_3$ the lines AA_1 and AA_2 are the asymptotic tangents at A. Let us put

$$l_1 = [AA_1], \qquad l_2 = [AA_2].$$

The asymptotic tangents l_1', l_2' at a neighboring point A' of A are given by the Taylor's expansions

$$l_i' = [AA_i] + [A\,dA_i] + [dA A_i] + \tfrac{1}{2}\{[A\,d^2A_i] + 2[dA\,dA_i] + [d^2A, A_i]\} + \cdots \qquad i = 1, 2.$$

It follows that

$$\begin{cases} Sl_1l_1' = (\omega^2\omega_1^3 - \omega^3\omega_1^2) + (3), \\ Sl_2l_2' = (\omega^3\omega_2^1 - \omega^1\omega_2^3) + (3), \\ Sl_1l_2' = -\omega^3 + (\omega^2\omega_2^3 - \omega^3\omega_2^2) - \tfrac{1}{2}\{d\omega^3 + \omega^3(\omega_0^0 + \omega_3^3) + \omega^1\omega_1^3 + \omega^2\omega_2^3\} + (3), \\ Sl_2l_1' = \omega^3 + (\omega^3\omega_1^1 - \omega^1\omega_1^3) + \tfrac{1}{2}\{d\omega^3 + \omega^3(\omega_0^0 + \omega_3^3) + \omega^1\omega_1^3 + \omega^2\omega_2^3\} + (3), \end{cases}$$

where (3) denotes a sum of terms of at least the third order. A line $\lambda_1l_1 + \lambda_2l_2$ through A on π intersects l_1' if and only if

$$S(\lambda_1l_1 + \lambda_2l_2)l_1' = 0,$$

i.e.

$$\frac{\lambda_1}{\lambda_2} = -\frac{Sl_2\,l_1'}{Sl_1\,l_1'}.$$

[5] G. Fubini et Čech, *Géométrie projective différentialle des surfaces,* Paris, 1931, pp. 4–6.

Similarly, a line $\rho_1 l_1 + \rho_2 l_2$ through A on π intersects l_2' if and only if

$$\frac{\rho_1}{\rho_2} = -\frac{Sl_2 l_2'}{Sl_1 l_2'}.$$

The four lines l_1, l_2, l_1', l_2' determine four points on the line of intersection of the tangent planes at A and A'. The cross ratio H of these four points is found to be

$$\frac{\rho_1}{\rho_2}\cdot\frac{\lambda_2}{\lambda_1} = \frac{(\omega^2\omega_1^3 - \omega^3\omega_1^2)(\omega^1\omega_2^3 - \omega^3\omega_2^1)}{(\omega^3)^2} = d\sigma^2.$$

Hence the projective linear element can be interpreted as follows:

Let l_1 and l_2 be the asymptotic tangents at A, and l_1', l_2' be those at a neighboring point A' of A not on π. Then the four lines l_1, l_2, l_1', l_2' intersect the line of intersection of the tangent planes π, π' at A, A' in four points whose cross ratio is equal to $d\sigma^2$.

4. The main theorem and two lemmas. Two non-holonomic surfaces which are projectively equivalent have the same projective linear element. It is easily verified that this remains true if they are related by a correlation. Our main theorem asserts that the converse also holds. We shall formulate it as follows:

THEOREM. *Suppose S and $\bar S$ be two non-holonomic surfaces whose tangent planes depend on three essential parameters. If there is a correspondence between S and $\bar S$ such that their projective linear elements are equal, then there exists a collineation or a correlation carrying S to $\bar S$.*

Before proving the theorem, we shall establish two lemmas.

LEMMA 1. *Let S and $\bar S$ be two non-holonomic surfaces having the same projective linear element and $\bar R$ be a family of canonical frames $\bar A\bar A_1\bar A_2\bar A_3$ of $\bar S$ such that at each point there is one and only one frame of $\bar R$. Then there exists a family R of canonical frames $AA_1A_2A_3$ of S having the same property and satisfying either the conditions*

$$(14)\quad \begin{cases} \omega^1 = \bar\omega^1, \quad \omega^2 = \bar\omega^2, \quad \omega^3 = \bar\omega^3, \quad \omega_1^3 = \bar\omega_1^3, \quad \omega_2^3 = \bar\omega_2^3, \\ \omega_1^2 = \bar\omega_1^2, \quad \omega_2^1 = \bar\omega_2^1, \quad \omega_0^0 - \omega_3^3 = \bar\omega_0^0 - \bar\omega_3^3, \end{cases}$$

or the conditions

$$(14')\quad \begin{cases} \omega^1 = -\bar\omega_2^3, \quad \omega^2 = -\bar\omega_1^3, \quad \omega^3 = -\bar\omega^3, \quad \omega_1^3 = -\bar\omega^2, \quad \omega_2^3 = -\bar\omega^1, \\ \omega_1^2 = -\bar\omega_1^2, \quad \omega_2^1 = -\bar\omega_2^1, \quad \omega_0^0 - \omega_3^3 = \bar\omega_0^0 - \bar\omega_3^3, \end{cases}$$

where ω_i^j and $\bar\omega_i^j$ $(i, j = 0, 1, 2, 3)$ are the Pfaffian forms corresponding to the two families R and $\bar R$ of frames respectively.

PROOF: We first take R to be the family of all the canonical frames attached to different points of the space, and let ω_i^j $(i, j = 0, 1, 2, 3)$ be the Pfaffian forms

in (2) satisfied by frames of this family. From the identity of the projective linear elements

$$(15)\qquad \frac{(\omega^2\omega_1^3 - \omega^3\omega_1^2)(\omega^1\omega_2^3 - \omega^3\omega_2^1)}{(\omega^3)^2} = \frac{(\bar{\omega}^2\bar{\omega}_1^3 - \bar{\omega}^3\bar{\omega}_1^2)(\bar{\omega}^1\bar{\omega}_2^3 - \bar{\omega}^3\bar{\omega}_2^1)}{(\bar{\omega}^3)^2},$$

we have

$$\omega^3 = \rho\bar{\omega}^3.$$

The third equation of (12) shows that we can choose a sub-family of R such that

$$(16)\qquad \omega^3 = \bar{\omega}^3.$$

Making use of this equation, we have, from (15)

$$(17)\qquad (\omega^2\omega_1^3 - \omega^3\omega_1^2)(\omega^1\omega_2^3 - \omega^3\omega_2^1) = (\bar{\omega}^2\bar{\omega}_1^3 - \omega^3\bar{\omega}_1^2)(\bar{\omega}^1\bar{\omega}_2^3 - \omega^3\bar{\omega}_2^1)$$

from which it follows

$$(18)\qquad (\omega^1\omega^2)^2 = (\bar{\omega}^1\bar{\omega}^2)^2$$

so that $\bar{\omega}^1$, $\bar{\omega}^2$ are multiples of ω^1 or ω^2. On the other hand, forming the bilinear covariant of (16), we get

$$[(\omega_0^0 - \omega_3^3 - \bar{\omega}_0^0 + \bar{\omega}_3^3)\omega^3] + \left(b - \frac{1}{b}\right)[\omega^1\omega^2] - \left(\bar{b} - \frac{1}{\bar{b}}\right)[\bar{\omega}^1\bar{\omega}^2] = 0$$

where $\bar{b}$ is defined by

$$\bar{\omega}_1^3 = \bar{b}\bar{\omega}^2.$$

The above equation holds, only when

$$(19)\qquad \omega_0^0 - \omega_3^3 - (\bar{\omega}_0^0 - \bar{\omega}_3^3) = \lambda\omega^3, \qquad \left(b - \frac{1}{b}\right)[\omega^1\omega^2] - \left(\bar{b} - \frac{1}{\bar{b}}\right)[\bar{\omega}^1\bar{\omega}^2] = 0.$$

From (12) and the relation

$$\delta(\omega_0^0 - \omega_3^3) = -2e_3^0\omega^3,$$

we see that we can make

$$(20)\qquad \omega^1 = \bar{\omega}^1, \qquad \omega_0^0 - \omega_3^3 = \bar{\omega}_0^0 - \bar{\omega}_3^3.$$

When these conditions hold, the family R of frames is such that to each point in space there is attached one and only one frame of R. In other words, there is a one-to-one correspondence between the frames of R and $\bar{R}$.

From equations (18) and (20) we get

$$\omega^2 = \pm\bar{\omega}^2.$$

According as the upper or the lower sign holds, we shall divide the discussion into two cases:

CASE 1. $\omega^2 = \bar{\omega}^2$. Then we have

$$(b - \bar{b})(b\bar{b} + 1) = 0.$$

If $b = \bar{b}$, equations (14) follow as a consequence of (17) and the equations already obtained. If $b\bar{b} + 1 = 0$, we change the family of frames R to the family R^* of frames $A^*A_1^*A_2^*A_3^*$, related to $AA_1A_2A_3$ as follows:

$$A^* = A, \qquad A_1^* = \frac{1}{b}A_1, \qquad A_2^* = bA_2, \qquad A_3^* = -A_3.$$

The two families R^* and $\bar{R}$ then satisfy the relations (14′).

CASE 2. $\omega^2 = -\bar{\omega}^2$. We have then $(b + \bar{b})(b\bar{b} - 1) = 0$. For the case $b + \bar{b} = 0$, equation (17), together with the change of frame

$$A^* = A, \qquad A_1^* = A_1, \qquad A_2^* = -A_2, \qquad A_3^* = A_3$$

gives (14), while for the case $b\bar{b} - 1 = 0$, equation (17), together with the change of frame

$$A^* = A, \qquad A_1^* = -\frac{1}{b}A_1, \qquad A_2^* = bA_2, \qquad A_3^* = -A_3$$

gives (14′). Thus the lemma is proved.

LEMMA 2. *Suppose there be two families of canonical frames $AA_1A_2A_3$, $\bar{A}\bar{A}_1\bar{A}_2\bar{A}_3$ of two non-holonomic surfaces S and $\bar{S}$ respectively. If the relations* (14) *hold, then we have*

$$\begin{cases} \omega_0^0 = \bar{\omega}_0^0, & \omega_1^1 = \bar{\omega}_1^1, \quad \omega_2^2 = \bar{\omega}_2^2, \quad \omega_3^3 = \bar{\omega}_3^3, \quad \omega_1^0 = \bar{\omega}_1^0, \\ \omega_2^0 = \bar{\omega}_2^0, & \omega_3^1 = \bar{\omega}_3^1, \quad \omega_3^2 = \bar{\omega}_3^2, \quad \omega_3^0 = \bar{\omega}_3^0. \end{cases} \tag{21}$$

PROOF. Since the relations (14) hold, we take their bilinear covariants and obtain

$$\begin{cases} [(\omega_0^0 - \omega_1^1 - \bar{\omega}_0^0 + \bar{\omega}_1^1)\omega^1] + [(\bar{\omega}_3^1 - \omega_3^1)\omega^3] = 0, \\ [(\omega_0^0 - \omega_2^2 - \bar{\omega}_0^0 + \bar{\omega}_2^2)\omega^2] + [(\bar{\omega}_3^2 - \omega_3^2)\omega^3] = 0, \\ b[(\omega_1^1 - \omega_3^3 - \bar{\omega}_1^1 + \bar{\omega}_3^3)\omega^2] + [(\omega_1^0 - \bar{\omega}_1^0)\omega^3] = 0, \\ \frac{1}{b}[(\omega_2^2 - \omega_3^3 - \bar{\omega}_2^2 + \bar{\omega}_3^3)\omega^1] + [(\omega_2^0 - \bar{\omega}_2^0)\omega^3] = 0, \\ [(\omega_1^0 + \frac{1}{b}\omega_3^2 - \bar{\omega}_1^0 - \frac{1}{b}\bar{\omega}_3^2)\omega^1] \\ \qquad + [(\omega_2^0 + b\omega_3^1 - \bar{\omega}_2^0 - b\bar{\omega}_3^1)\omega^2] + 2[(\omega_3^0 - \bar{\omega}_3^0)\omega^3] = 0, \\ [(\omega_1^1 - \omega_2^2 - \bar{\omega}_1^1 + \bar{\omega}_2^2)(\alpha\omega^1 + \gamma\omega^3)] + [(\omega_1^0 - b\omega_3^2 - \bar{\omega}_1^0 + b\bar{\omega}_3^2)\omega^2] = 0, \\ [(\omega_2^2 - \omega_1^1 - \bar{\omega}_2^2 + \bar{\omega}_1^1)(\alpha^*\omega^2 + \gamma^*\omega^3)] + [(\omega_2^0 - \frac{1}{b}\omega_3^1 - \bar{\omega}_2^0 + \frac{1}{b}\bar{\omega}_3^1)\omega^1] = 0 \end{cases} \tag{22}$$

where b, α, α^*, γ, γ^* are given by (10). From (22) we can put

$$(23)\quad \begin{cases} \omega_0^0 - \omega_1^1 - \bar{\omega}_0^0 + \bar{\omega}_1^1 = r\omega^3, & \bar{\omega}_3^1 - \omega_3^1 = r\omega^1 + u\omega^3, \\ \omega_0^0 - \omega_2^2 - \bar{\omega}_0^0 + \bar{\omega}_2^2 = r^*\omega^3, & \bar{\omega}_3^2 - \omega_3^2 = r^*\omega^2 + u^*\omega^3, \\ \bar{\omega}_1^0 - \omega_1^0 = br\omega^2 + bu^*\omega^3, & \bar{\omega}_2^0 - \omega_2^0 = \dfrac{1}{b}\, r^*\omega^1 + \dfrac{1}{b}\, u\omega^3, \\ \bar{\omega}_3^0 - \omega_3^0 = \frac{1}{2}\left(b + \dfrac{1}{b}\right)(u^*\omega^1 + u\omega^2) + \frac{1}{2}v\omega^3. \end{cases}$$

Taking the bilinear covariant of the first two equations of (23), we obtain

$$(24)\quad \begin{cases} 2[(\bar{\omega}_1^0 - \omega_1^0)\omega^1] + [(\bar{\omega}_2^0 - b\bar{\omega}_3^1 - \omega_2^0 + b\omega_3^1)\omega^2] + [(\bar{\omega}_3^0 - \omega_3^0)\omega^3] \\ \qquad = [\{dr + r(\omega_0^0 - \omega_3^3)\}\omega^3] + r\left(b - \dfrac{1}{b}\right)[\omega^1\omega^2], \\ 2[(\bar{\omega}_2^0 - \omega_2^0)\omega^2] + \left[\left(\bar{\omega}_1^0 - \dfrac{1}{b}\bar{\omega}_3^2 - \omega_1^0 + \dfrac{1}{b}\omega_3^2\right)\omega^1\right] + [(\bar{\omega}_3^0 - \omega_3^0)\omega^3] \\ \qquad = [\{dr^* + r^*(\omega_0^0 - \omega_3^3)\}\omega^3] + r^*\left(b - \dfrac{1}{b}\right)[\omega^1\omega^2]. \end{cases}$$

A comparison of the coefficients of $[\omega^1\omega^2]$ on both sides of (24) gives

$$(25)\qquad r^* = (4b^2 - 1)r, \qquad r = \frac{4 - b^2}{b^2}\, r^*$$

from which we get

$$(26)\qquad (b^4 - 4b^2 + 1)r = 0.$$

Suppose that $b^4 - 4b^2 + 1 \neq 0$. It follows from (26) and (24) that

$$r = r^* = u = u^* = 0.$$

The relations (23) then become

$$\omega_0^0 - \omega_1^1 = \bar{\omega}_0^0 - \bar{\omega}_1^1, \quad \omega_3^1 = \bar{\omega}_3^1, \quad \omega_1^0 = \bar{\omega}_1^0, \quad \bar{\omega}_3^0 - \omega_3^0 = \tfrac{1}{2}v\omega^3,$$
$$\omega_0^0 - \omega_2^2 = \bar{\omega}_0^0 - \bar{\omega}_2^2, \quad \omega_3^2 = \bar{\omega}_3^2, \quad \omega_2^0 = \bar{\omega}_2^0$$

from which, by forming the bilinear covariants, we get $v = 0$. If this is the case, the lemma is established. If $b^4 - 4b^2 + 1 = 0$, while α and α^* do not both vanish, we still have, on account of (25) and the last two equations of (22),

$$r = r^* = 0.$$

From these equations, the lemma can be established as above.

It remains for us to consider the case

$$(27)\qquad \alpha = \alpha^* = b^4 - 4b^2 + 1 = 0.$$

Let

$$(28)\qquad \begin{cases} dr = r_1\omega^1 + r_2\omega^2 + r_3\omega^3, & dr^* = r_1^*\omega^1 + r_2^*\omega^2 + r_3^*\omega^3, \\ \omega_0^0 - \omega_3^3 = \alpha_1\omega^1 + \alpha_2\omega^2 + \alpha_3\omega^3. \end{cases}$$

Then equations (24) give

$$(29)\qquad \begin{cases} \frac{1}{2}u^*\left(\frac{1}{b} - 3b\right) = r_1 + r\alpha_1, & \frac{1}{2}u\left(b - \frac{3}{b}\right) = r_2^* + r^*\alpha_2, \\ \frac{1}{2}u^*\left(\frac{3}{b} - b\right) = r_1^* + r^*\alpha_1, & \frac{1}{2}u\left(3b - \frac{1}{b}\right) = r_2 + r\alpha_2, \end{cases}$$

from which we have, by making use of (24),

$$u = u^* = 0.$$

Then equations (23) take the following form:

$$(30)\qquad \begin{cases} \omega_0^0 - \omega_1^1 - \bar{\omega}_0^0 + \bar{\omega}_1^1 = r\omega^3, & \bar{\omega}_3^1 - \omega_3^1 = r\omega^1, & \bar{\omega}_1^0 - \omega_1^0 = br\omega^2, \\ \omega_0^0 - \omega_2^2 - \bar{\omega}_0^0 + \bar{\omega}_2^2 = r^*\omega^3, & \bar{\omega}_3^2 - \omega_3^2 = r^*\omega^2, & \bar{\omega}_2^0 - \omega_2^0 = \frac{1}{b}\, r^*\omega^1, \\ \bar{\omega}_3^0 - \omega_3^0 = \frac{1}{2}v\omega^3. \end{cases}$$

Now the Pfaffian forms ω_i^j are not arbitrary. In fact, by taking the bilinear covariants of the first two equations of (10), we get

$$\gamma\left(b + \frac{1}{b}\right)[\omega^3\omega^1] + b[(\omega_1^1 + \omega_2^2 - \omega_0^0 - \omega_3^3)\omega^2] + [(\omega_1^0 + b\omega_3^2)\omega^3] = 0,$$

$$\gamma^*\left(b + \frac{1}{b}\right)[\omega^3\omega^2] + \frac{1}{b}\,[(\omega_1^1 + \omega_2^2 - \omega_0^0 - \omega_3^3]\omega^1] + \left[\left(\omega_2^0 + \frac{1}{b}\,\omega_3^1\right)\omega^3\right] = 0.$$

Thus we can write

$$(31)\qquad \begin{cases} \omega_1^0 + b\omega_3^2 = \gamma\left(b + \frac{1}{b}\right)\omega^1 + (\cdots)\omega^2 + (\cdots)\omega^3, \\ \omega_2^0 + \frac{1}{b}\,\omega_3^1 = \gamma^*\left(b + \frac{1}{b}\right)\omega^2 + (\cdots)\omega^1 + (\cdots)\omega^3. \end{cases}$$

Two cases must now be considered separately—the case $\gamma = \gamma^* = 0$ and the case by which γ and γ^* do not both vanish. In the former case, we get $r = r^* = v = 0$, the derivation of which involves a long calculation, whose details we omit here. In the latter case, we assume, for definiteness, $\gamma \neq 0$. By forming the bilinear covariant of the identity

$$r^*(\bar{\omega}_1^0 - \omega_1^0) - br(\bar{\omega}_3^2 - \omega_3^2) \equiv 0$$

we obtain

$$\gamma\left\{\left(b + \frac{1}{b}\right)rr^* + \left(br^2 + \frac{1}{b}\,r^{*2}\right)\right\}[\omega^1\omega^3] + (\cdots)[\omega^2\omega^3] + (\cdots)[\omega^1\omega^2] = 0.$$

This equation, together with equations (25) and (27), also gives

$$r = r^* = v = 0.$$

Our proof is therefore complete.

PROOF OF THE MAIN THEOREM. Let S and $\bar{S}$ be two non-holonomic surfaces having the same projective linear element. From lemma 1, there are two families of canonical frames of S and $\bar{S}$ respectively such that either the conditions (14) or the conditions (14′) hold. In the former case, lemma 2 shows that S can be transformed to $\bar{S}$ by a collineation. In the latter case, we introduce plane coordinates by putting

$$a = -[AA_1A_2], \qquad a_1 = [AA_1A_3], \qquad a_2 = -[AA_2A_3], \qquad a_3 = [A_1A_2A_3].$$

Then a, a_1, a_2, a_3 satisfy the equations[6]

$$\begin{cases} da = -\omega_3^3 a - \omega_2^3 a_1 - \omega_1^3 a_2 - \omega^3 a_3, \\ da_1 = -\omega_3^2 a - \omega_2^2 a_1 - \omega_1^2 a_2 - \omega^2 a_3, \\ da_2 = -\omega_3^1 a - \omega_2^1 a_1 - \omega_1^1 a_2 - \omega^1 a_3, \\ da_3 = -\omega_3^0 a - \omega_2^0 a_1 - \omega_1^0 a_2 - \omega^0 a_3 \end{cases}$$

and lemma 2 gives

$$\begin{cases} -\omega_3^3 = \bar{\omega}_0^0, \quad -\omega_3^2 = \bar{\omega}_1^0, \quad -\omega_3^1 = \bar{\omega}_2^0, \quad -\omega_3^0 = \bar{\omega}_3^0, \quad -\omega_2^2 = \bar{\omega}_1^1 \\ -\omega_2^0 = \bar{\omega}_3^1, \quad -\omega_1^1 = \bar{\omega}_2^2, \quad -\omega_1^0 = \bar{\omega}_3^2, \quad -\omega_0^0 = \bar{\omega}_3^3, \end{cases}$$

which proves that the non-holonomic surfaces S and $\bar{S}$ are related by a correlation. Hence our main theorem is proved.

TSING HUA UNIVERSITY,
KUNMING, CHINA.

[6] G. FUBINI ET ČECH, ibid. p. 220.

ANNALS OF MATHEMATICS
Vol. 44, No. 4, October, 1943

ON THE IMPORTANCE OF THE RELATION $[(A, B), (A, C)] < (A, [(B, C), (C, A), (A, B)]$ BETWEEN THREE ELEMENTS OF A STRUCTURE

BY HENRY LÖWIG

(Received January 8, 1940)

If A, B, and C are elements of an arbitrary structure then we have always

$$(1) \qquad [(A, B), (A, C)] \leqq (A, [(B, C), (C, A), (A, B)]).$$

(We will use in this paper the terminology and notation introduced in the paper (II); the marks $<$ and $>$ shall exclude the equality.) If the structure is especially a so-called Dedekind structure then in (1) the equality holds always i.e. the equation

$$(2) \qquad [(A, B), (A, C)] = (A, [(B, C), (C, A), (A, B)])$$

is generally valid. (See for instance (II), p. 413, equation (4).) In this paper we inquire whether conversely every structure in which the equation (2) is generally valid is a Dedekind structure, and how those structures in which (2) is not generally valid are characterized.

THEOREM 1. *If A, B, and C are elements of a structure, and between the cross-cuts (B, C), (C, A), and (A, B) there is at least one relation of inclusion then* (2) *is valid.*

PROOF. If for instance $(B, C) \leqq (C, A)$, then we have obviously $(B, C) \leqq (A, B)$ too. Therefore it is sufficient to consider the two cases $(B, C) \leqq (C, A)$ and $(C, A) \leqq (B, C)$.

1. Let (B, C) be $\leqq (C, A)$. In this case we have

$$[(B, C), (C, A), (A, B)] = [(C, A), (A, B)];$$

from this equation ensues immediately the asserted equation (2).

2. Let (C, A) be $\leqq (B, C)$. In this case we have obviously

$$[(B, C), (C, A), (A, B)] \leqq B$$

whence

$$(A, [(B, C), (C, A), (A, B)]) \leqq (A, B),$$

and consequently

$$(A, [(B, C), (C, A), (A, B)]) \leqq [(A, B), (A, C)].$$

But the last inequality implies together with the inequality (1) the asserted equation (2).

THEOREM 2. *If between the elements A, B, and C of a structure there is at least one relation of inclusion then* (2) *is valid.*

Theorem 2 ensues immediately from Theorem 1:

There are non-Dedekind structures in which nevertheless (2) is satisfied for arbitrary elements A, B, and C. We have for instance the

THEOREM 3: *In any non-Dedekind structure of the fifth order* (2) *is generally valid.*

PROOF. If Σ is a non-Dedekind structure of the fifth order then we can (see (I), §6) denote its elements by P, Q, R, S, and T in such a way that the following inequalities hold:

$$(3) \qquad \begin{cases} P < Q, P < R, P < S, P < T; \\ Q < T; \\ R < S, R < T; \\ S < T. \end{cases}$$

From (3) it is directly evident that between any three elements of Σ there is always at least one relation of inclusion. Now it ensues immediately from Theorem 2 that Theorem 3 is true.

Theorem 3 will be still further generalized in the following.

THEOREM 4. *There are structures in which* (2) *is not generally valid.*

PROOF. We consider a set Λ consisting of 9 elements E, L, M, N, X_1, Y, Z, X, F. Now we will establish between these elements of Λ the following 24 proper relations of inclusion:

$$(4) \qquad \begin{cases} E < L, E < M, E < N, E < X_1, E < Y, E < Z, E < F; \\ L < Y, L < Z, L < F; \\ M < Z, M < X_1, M < X, M < F; \\ N < X_1, N < X, N < Y, N < F; \\ X_1 < X, X_1 < F; \\ Y < F; \\ Z < F; \\ X < F. \end{cases}$$

One persuades himself easily that by the establishments (4) Λ becomes a partly ordered set. (See also figure on p. 575.) Moreover this partly ordered set is even a structure, and cross-cut and union of any two distinct elements of this structure can be found from the table on p. 576.

According to this table we have

$$[(X, Y), (X, Z)] = [N, M] = X_1,$$

and

$$(X, [(Y, Z), (Z, X), (X, Y)]) = (X, [L, M, N]) = (X, [Z, N]) = (X, F) = X.$$

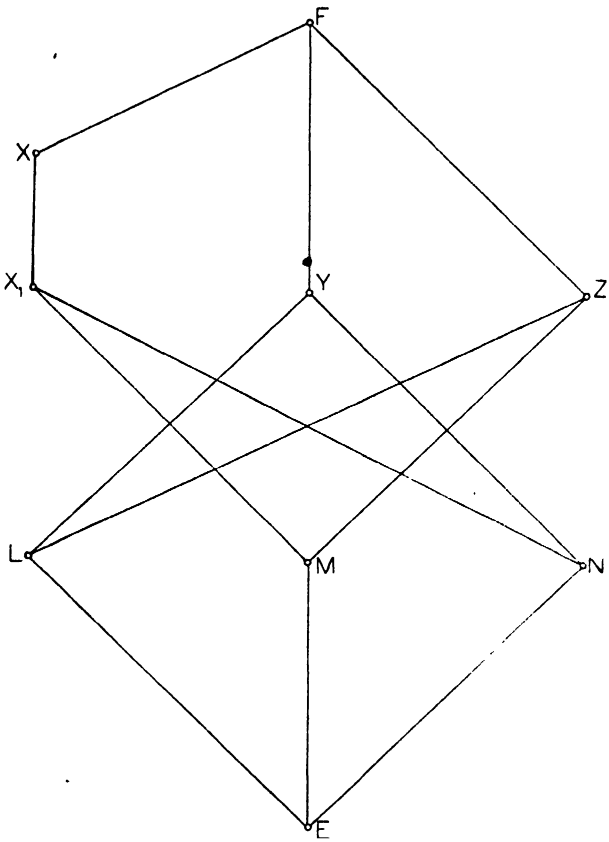
Hence

$$[(X, Y), (X, Z)] \neq (X, [(Y, Z), (Z, X), (X, Y)]).$$

In the structure Λ considered here the equation (2) is consequently not generally valid. (On the other hand one can easily persuade himself that the triples

X, Y, Z and X, Z, Y are the only triples of elements of Λ which do not satisfy the equation (2).)

In the following the marks E, L, M, N, X_1, Y, Z, X, F, and Λ shall have always the significance stated.



THEOREM 5. *If A, B, and C are elements of an arbitrary structure Σ then the set Λ' of the 9 elements E', L', M', N', X_1', Y', Z', X', F' defined by the equations*

$$(5)\qquad \begin{cases} E' = (A, B, C) \\ L' = (B, C) \\ M' = (C, A) \\ N' = (A, B) \\ X_1' = [(C, A), (A, B)] \\ Y' = [(A, B), (B, C)] \\ Z' = [(B, C), (C, A)] \\ X' = (A, [(B, C), (C, A), (A, B)]) \\ F' = [(B, C), (C, A), (A, B)] \end{cases}$$

is a substructure of Σ, *and if we assign to every element of* Λ *the corresponding accented element of* Λ', *then this correspondence is a homomorphism between* Λ *and* Λ'.

To prove Theorem 5 we have to demonstrate that the table below continues to be valid if we replace all elements appearing in it by the corresponding

	E	L	M	N	X_1	Y	Z	X	F	[ ]
E		L	M	N	X_1	Y	Z	X	F	E
L	E		Z	Y	F	Y	Z	F	F	L
M	E	E		X_1	X_1	F	Z	X	F	M
N	E	E	E		X_1	Y	F	X	F	N
X_1	E	E	M	N		F	F	X	F	X_1
Y	E	L	E	N	N		F	F	F	Y
Z	E	L	M	E	M	L		F	F	Z
X	E	E	M	N	X_1	N	M		F	X
F	E	L	M	N	X_1	Y	Z	X		F
( )	E	L	M	N	X_1	Y	Z	X	F	

accented elements. As this demonstration involves no difficulties we will content ourselves to prove the equation $[X', Y'] = F'$.

We have obviously $X' \geqq (C, A)$. Hence

$$[X', Y'] \geqq [(B, C), (C, A), (A, B)]$$

or

$$[X', Y'] \geqq F'. \tag{6}$$

On the other side we have $X' \leqq F'$ and $Y' \leqq F'$, and consequently also

$$[X', Y'] \leqq F'. \tag{7}$$

(6) and (7) yield the equation

$$[X', Y'] = F'$$

we have wanted to prove.

THEOREM 6. (Lemma.) *Let to every element U of a structure Σ be assigned an element U' of another structure Σ', where this correspondence is a homomorphism between Σ and Σ'. If $U < V$ implies $U' < V'$ then the given homomorphism is an isomorphism.*

PROOF. If $U \neq V$, and between U and V there exists a relation of inclusion then we have, according to the supposition, also $U' \neq V'$. But if we have $U \neq V$ without a relation of inclusion existing between U and V then we have certainly

$$(U, V) < U.$$

Therefore we have, according to what we have supposed in Theorem 6, also

$$(U', V') < U'.$$

The last inequality implies that also in this case $U' \neq V'$. Hence to distinct elements of Σ correspond always distinct elements of Σ'.

THEOREM 7. *If A, B, and C are elements of a structure Σ which do not satisfy the equation (2) then the homomorphism between Λ and Λ' existing according to Theorem 5 is an isomorphism.*

PROOF. According to Theorem 6 it is sufficient to demonstrate that none of the inequalities (4) turns into an equality if we replace both its sides by the corresponding accented elements. Moreover we can restrict ourselves to prove the 13 inequalities

$$(8)\qquad \begin{cases} E' < L', \quad E' < M', \quad E' < N', \\ L' < Y', \quad L' < Z', \\ M' < Z', \quad M' < X_1', \\ N' < X_1', \quad N' < Y', \\ X_1' < X', \\ Y' < F', \\ Z' < F', \\ X' < F', \end{cases}$$

because the other inequalities coming into question, namely the inequalities

$$(9)\qquad \begin{cases} E' < X_1', \quad E' < Y', \quad E' < Z', \quad E' < X', \quad E' < F', \\ L' < F', \\ M' < X', \quad M' < F', \\ N' < X', \quad N' < F', \\ X_1' < F' \end{cases}$$

ensue from the inequalities (8). The discussion shall be divided into 4 steps.

1. If the equation $E' = L'$ were valid we should have

$$(A, B, C) = (B, C)$$

and consequently

$$(B, C) \leqq (A, B);$$

hence, according to Theorem 1, the elements A, B, and C would satisfy the equation (2), contrarily to the supposition of Theorem 7. In corresponding way we can refute the hypotheses $E' = M'$ and $E' = N'$.

2. $L' = Y'$ or $(B, C) = [(A, B), (B, C)]$ would imply $(A, B) \leqq (B, C)$. Again it follows from Theorem 1 that the hypothesis is absurd. Similarly also the hypotheses $L' = Z'$, $M' = Z'$, $M' = X_1'$, $N' = X_1'$, and $N' = Y'$ yield contradictions.

3. According to the supposition of Theorem 7 we have $X_1' \neq X'$.

4. If the equation $X' = F'$ were valid then we should have

$$(L', X') = (L', F'),$$

and consequently

$$E' = L'.$$

This would contradict what we have proved under 1. Just so from the hypothesis $Y' = F'$ we can derive the equation $E' = M'$, from the hypothesis $Z' = F'$ the equation $E' = N'$, and by this a contradiction to what we have proved under 1.

Now the following Theorems 8, 9, and 10 are evident.

THEOREM 8. *Any structure in which the equation* (2) *is not generally valid contains at least one sub-structure of the ninth order having the same property.*

THEOREM 9. *Any structure of the ninth order in which the equation* (2) *is not generally valid is isomorphic with the structure* Λ.

THEOREM 10. *In any structure of at most eighth order the equation* (2) *holds for arbitrary elements* A, B, *and* C.

Theorem 10 represents the announced generalization of Theorem 3.

It is evident that we could investigate the equation

$$([A, B], [A, C]) = [A, ([B, C], [C, A], [A, B])] \tag{10}$$

dually corresponding to the equation (2) exactly in the same way as we have just explored the equation (2). Yet, while the Dedekind axiom

$$[(A, [B, C]), (B, C)] = ([A, (B, C)], [B, C])$$

is identical with its dual counterpart the corresponding assertion about the equation (2) is not true.

THEOREM 11. *In our structure* Λ *the equation* (10) *is generally valid.*

The proof follows easily from the dual counterpart of Theorem 9. Of course Theorem 11 can also be verified simply by the table on p. 576.

THEOREM 12. *The assertion that in a structure the equation* (2) *holds for arbitrary*

elements, A, B, and C, and the assertion dually corresponding to this assertion are not equivalent.

PROOF. See Theorem 11.

PRAGUE

References

(I) Richard Dedekind, *Über die von drei Moduln erzeugte Dualgruppe.* Mathematische Annalen 53(1900), p. 371–403.

(II) Oystein Ore, *On the foundation of abstract algebra.* I. Annals of Mathematics 36(1935), p. 406–437.

ANNALS OF MATHEMATICS
Vol. 44, No. 4, October, 1943

ON THE THEORY OF PARTIALLY ORDERED LINEAR SYSTEMS AND LINEAR SPACES

BY DIMITRY WASSILKOFF

(Received February 8, 1943)

In the present work we consider some problems concerning the theory of partially ordered linear systems and linear spaces, which have been developed by F. Riesz, H. Freudenthal, L. Kantorovitch, M. Krein, Sh. Kakutani and others. The main purpose of this paper is to reveal the rôle of a group of axioms analogous to the so called Axiom of Archimedes (or Eudoxe) for real numbers. The strongest of these axioms—Axiom A—determines a class of partially ordered linear systems which proves to be rather important in various aspects of the theory.

§1 contains the complete list of notions and axioms which we shall deal with in the following. As to the elementary properties of the linear system connected with the ordering[1] we refer chiefly to H. Freudenthal [1] and L. Kantorovitch [2].

§1. Notations and definitions

Let $E = \{x, y, z, \cdots\}$ be a *linear system* (an Abelian group with real numbers as operators [3]) and a *partially ordered set*, i.e. between some pairs of elements x, y, of E ($x \neq y$) a relation $x < y$ is defined which satisfies the usual conditions:

1. $x < y$ and $y < x$ are inconsistent;
2. $x < y$ and $y < z$ imply $x < z$.

We shall use the following notations:

$x > y$ denotes that $y < x$.

$x \leqq y$ (or $y \geqq x$) denotes that either $x < y$ or $x = y$.

$x \| y$ denotes that $x \neq y$ and neither $x < y$ nor $x > y$ takes place; we say that x and y are *incomparable*.

$x \in M$ denotes that x is an element of the set M.

$x \bar{\in} M$ denotes that the set M does not contain x.

$M_1 \subset M_2$ (or $M_2 \supset M_1$) denotes that M_1 is a subset of M_2 (M_1 and M_2 may coincide).

$M_1 \leqq M_2$ (or $M_2 \geqq M_1$) denotes that $x \leqq y$, whenever $x \in M_1$, $y \in M_2$.

$N(M)$ denotes the set of all $x \leqq \mathfrak{M}$.

$H(M)$ denotes the set of all $y \geqq \mathfrak{M}$.

We introduce the notation $E^+(E^-)$ for the set $H(O)$ (resp. $N(O)$), where O is the null-element of the linear system E. Every element of E^+ (resp. E^-) except O is called *positive* (resp. *negative*).

Every element $x \in H(M)$ is called an *upper bound*, abbreviated: u.b., of the set M. Every $x \in N(M)$ is called a *lower bound*, abbreviated: l.b., of $\mathfrak{M}$.

The set M is *bounded* if there exist an u.b. as well as a l.b. of M.

---

[1] We always say "ordering" instead of "partial ordering."

The element x is the *least upper bound*, abbreviated: l.u.b., of the set M if $x \in H(M)$ and $x \leqq H(M)$.

The element $x \in N(M)$ such that $x \geqq N(M)$ is called the *greatest lower bound*, abbreviated: g.l.b., of M.

The l.u.b. (g.l.b.) of the set M we denote by sup M (resp. inf. M). If M contains only a finite number of elements $x_1, x_2, \cdots, x_r$, then we use the notations

$$\sup M = x_1 \vee x_2 \vee \cdots \vee x_r .$$

$$\inf M = x_1 \wedge x_2 \wedge \cdots \wedge x_r .$$

$M_1 \cup M_2$ ($M_1 \cap M_2$) denotes the union (resp. the intersection) of the sets M_1 and M_2.

$\bigcup_\alpha M_\alpha$ ($\bigcap_\alpha M_\alpha$) denotes the union (resp. the intersection) of the system of sets $\{M_\alpha\}$.

$M_1 \dot{-} M_2$ denotes the set of such x that $x \in M_1$, $x \bar{\in} M_2$.

$M_1 \pm M_2$, where M_1, M_2 are subsets of a linear system, denotes the set of the elements $x \pm y$, where x and y range through M_1 and M_2 respectively.

λM, where M is a subset of a linear system and λ is a real number, denotes the set of elements λx, where $x \in M$.

If x is an element of a linear system then D_x denotes the set of elements λx, where $-\infty < \lambda < +\infty$.

The subset of D_x consisting of λx with $0 < \lambda < +\infty$ is denoted by R_x.

If M is an arbitrary subset of a linear system E then $L(M)$ denotes the *linear hull* of M, i.e. the minimal linear subsystem $E' \subset E$ which contains M.

Return to our system E. The following axioms are always supposed to be satisfied:

AXIOM I. *$x < y$ implies $x + z < y + z$ for every $z \in E$.*

AXIOM II. *If $x < y$ then for every number $\lambda > 0$: $\lambda x < \lambda y$.*

AXIOM III. *For any two elements $x \in E$, $y \in E$ there exists a $z \in E$ such that $z \geqq x$, $z \geqq y$.*

The system E which satisfies Axioms I, II and III is called a *partially ordered linear system.*

Let E' be a linear subsystem of E. E' itself necessarily satisfies Axioms I and II (it is possible, of course, that every two elements of E' are incomparable). As to Axiom III it may be fulfilled or not. We say that E' is a *proper subsystem* of E if E' satisfies Axiom III.

Furthermore we shall use some of the following axioms:

AXIOM H*. *For every $x > 0$ the set Rx has no u.b.'s.*

AXIOM H. *For every $x \in E$ the set D_x is not bounded.*

AXIOM R. *If for a given $x \parallel 0$ the set R_x has an u.b. then there exists also a l.b. of R_x.*

We say, for instance, that the system E is E_H ($E_{H,R}$) if E satisfies Axiom H (resp. Axioms H and R).

It is easy to see that every E_H is E_{H^*}. The system E_{H^*} that is not E_H we call *weakly homogeneous*. The system E_H we call *homogeneous*. We say that E is a *regular* system if E is E_R. If E is weakly homogeneous and regular then we call it *weakly Archimedean*. If E is $E_{H,R}$ then we call it *Archimedean*.

It is easy to see that the weakly Archimedean and Archimedean systems can be characterized respectively by the following axioms:

AXIOM A*. *If for a given $x \in E$ the set R_x has an u. b. then necessarily $x \leqq 0$ or $x \parallel 0$; in the latter case R_x has also a l.b.*

AXIOM A. *If for a given $x \in E$ the set R_x has an u. b. then necessarily $x \leqq 0$.*

We introduce also

AXIOM B*. *E contains such an element $u > 0$ that every $x \in E$ can be represented in the form $x = \lambda u - x'$, where $\lambda \geqq 0$, $x' \in E^+$.*

AXIOM B. *E is a Banach space the unit sphere of which has an u.b.*

AXIOM C*. *Every finite subset M of E has a l.u.b. and a g.l.b.*

AXIOM C. *Every subset M of E that has an u.b. has a l.u.b.*

The system E_{B^*} is said to have an *axial element* u. The system E_{C^*} is a *lattice* according to the terminology of G. Birkhoff [4]. We call E *closed* if E is E_C.[2]

REMARK 1.1. There is an essential difference between Axioms H^*, H, R, A^*, A and Axioms B^*, B, C^*, C. The axioms of the first group are *cogradient*, i.e. every proper subsystem $E' \subset E$ is, for instance, E'_H if E itself is E_H. The axioms of the second group do not possess this property.

REMARK 1.2. In Axioms H^*, R, A^* and A we can replace R_x by its subset $\{nx\}$, where $n = 1, 2, \cdots$. In fact, if

$$x_1 \leqq y,$$
$$x_2 \leqq y$$

and α, β are non-negative numbers with $\alpha + \beta = 1$ then

$$\alpha x_1 \leqq \alpha y$$
$$\beta x_2 \leqq \beta y,$$

whence

$$\alpha x_1 + \beta x_2 \leqq \alpha y + \beta y = (\alpha + \beta)y = y,$$

i.e., every $N(y)$ (as well as every $H(y)$) is convex. Suppose now that

$$nx \leqq y$$

for $n = 1, 2, \cdots$. By Axiom III y can be chosen in E^+, i.e., the inequality holds for $n = 0$ too. If $\lambda \geqq 0$ is arbitrary, then $\lambda = n + \alpha$, where n is a non-negative integer and $0 \leqq \alpha < 1$, and we have $\lambda x = (n + \alpha)x = [(1 - \alpha)n + \alpha(n + 1)]x = (1 - \alpha)nx + \alpha (n + 1)x$. Since

$$nx \leqq y,$$
$$(n + 1)x \leqq y,$$

[2] Such systems are usually called *complete*.

we obtain

$$\lambda x = (1 - \alpha)nx + \alpha(n + 1)x \leqq y.$$

§2. Positive linear functions

The real function $l(x)$ defined on E is called *linear* if for arbitrary $x_1 \in E$, $x_1 \in E$, $x_2 \in E$ and numbers λ_1, λ_2

$$l(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 l(x_1) + \lambda_2 l(x_2).$$

THEOREM 2.1. *Let E' be a linear subsystem of E, l' a linear function defined on E'. There exists a linear function l on E such that $l(x) = l'(x)$, whenever $x \in E'$.*

THEOREM 2.2. *Let E' be a linear subsystem of E. If $E' \neq E$ then there exists a linear function l such that $l \neq 0$ and $l(x) = 0$, whenever $x \in E'$.*

THEOREM 2.3. *For every $x_0 \in E$ there exists a linear function $l(x)$ on E such that $l(x_0) = 1$.*

These theorems are the analogues of the well known theorems in the theory of linear functionals in Banach spaces and can be proved in essentially the same way.

The following lemmas will be useful.

LEMMA 2.1. *Axiom III is equivalent (Axioms I and II being fulfilled) to each of the following conditions:*

(a) $E = L(E^+)$;

(b) *Every $x \in E$ can be represented $x = x' - x''$, where $x' \in E^+$, $x'' \in E^+$.*

(c) *If a linear function $l(x)$ vanishes identically on E^+ then $l = 0$.*

PROOF. Axiom III is evidently equivalent to (b). The equivalence of it to the other conditions can be proved according to the scheme

$$\text{(b)} \to \text{(a)} \to \text{(c)} \to \text{(b)}.$$

The first and the second steps are also obvious. Suppose (b) is not fulfilled. Then the linear subsystem E' consisting of the elements $x' - x''$, where $x' \in E^+$, $x'' \in E^+$, is a proper subsystem of E and does not coincide with E. Therefore there exists a linear function $l \neq 0$ such that $l(x) = 0$ for all $x \in E'$.

LEMMA 2.2. *If E is a finite-dimensional system then Axiom III is equivalent to the following condition:*

(d) *E^+ contains an inner point of E in the sense of the natural topology in E.*[3]

PROOF. Obvious, since in this case (d) is equivalent to (a) in Lemma 2.1.

The linear function $p(x)$ on E is called *positive* if $p \neq 0$ and $p(x) \geqq 0$, whenever $x \in E^+$.

THEOREM 2.4. *Let E be a weakly homogeneous system, E' an arbitrary proper*

---

[3] By the natural topology of the n-dimensional linear system E we understand the unique topology in E with respect to which the linear operations in E are continuous. It can be introduced by the norm: if $x = \sum_{i=1}^{n} \xi_i x_i$ then $\| x \| = \sum_{i=1}^{n} |\xi_i|$, where $\{x_1, x_2, \cdots, x_n\}$ is the basis of E.

subsystem of E, p' a positive linear function on E'. There exists then a positive linear function p on E such that $p(x) = p'(x)$ for every $x \in E'$.

PROOF. Consider the totality $\sum$ of all complex-valued functions $\sigma(x)$(x in E), with $|\sigma(x)| \equiv 1$. We introduce a *weak topology* in Σ defining the neighbourhoods in Σ as follows: a number $\epsilon < 0$ and a finite aggregate of elements $x_i \in E$ ($i = 1, 2, \cdots, r$) are fixed and the neighbourhood $U(\sigma_0; x_1, \cdots, x_r; \epsilon)$ of $\sigma_0(x)$ is defined as the set of such $\sigma(x)$ that

$$|\sigma(x_i) - \sigma_0(x_i)| < \epsilon \qquad (i = 1, 2, \cdots, r).$$

Let K denote the unit circle in the complex plain. Then Σ is the topological product (in the sense of A. Tychonoff) of the spaces K_x, where x runs over E, and every K_x is K,

$$\Sigma = \prod_x^E K_x, \qquad K_x \equiv K.$$

Since K is bicompact, Σ is also bicompact [5].

Σ contains the subspace Φ consisting of the functions $\varphi(x) = e^{il(x)}$, where $l(x)$ is an arbitrary linear function on E. It can be easily verified that Φ is closed in Σ, consequently Φ is bicompact itself.

Let x_0 be an arbitrary positive element of E. If $x_0 \bar{\in} E'$ then we denote by Π_{x_0} the set of the functions $\varphi(x) = e^{il(x)}$ with $l(x_0) \geqq 0$. If $x_0 \in E'$ then we denote by Π_{x_0} the set of such $\varphi(x) = e^{il(x)}$ that $l(x_0) = p'(x_0)$. It follows from Theorems 2.1 and 2.3 that $\Pi_x \neq 0$ for every $x > 0$. It is also easy to see that each Π_x is closed in Φ.

Consider an arbitrary finite system of the sets Π_x

$$\Pi_{x_1}, \Pi_{x_2}, \cdots, \Pi_{x_n} \qquad (x_i > 0; i = 1, 2, \cdots, n).$$

If E' does not contain any of the elements $x_1, x_2, \cdots, x)$, then the function $\varphi(x) \equiv 1$ belongs to every Π_{x_i} ($i = 1, 2, \cdots, n$). If some of x_i belong to E' then we may suppose that $\{x_1, x_2, \cdots, x_k\} \subset E'$ ($k \leqq n$). We take the finite-dimensional subsystems $E_1 = L(x_1, x_2, \cdots, x_n)$ and $E_2 = L(x_1, x_2, \cdots, x_k)$ of E.

There exists a positive linear function $p_1(x)$ on E_1 that coincides with $p'(x)$ on E_2. Assume the contrary: there would exist then an $x_0 \in E_1$, $x_0 > 0$, such that 1) necessarily $p_1(x_0) > 0$, 2) x_0 is the limit (in the sense of the natural topology in E_1) of a sequence of elements $y_m < 0$ ($y_m \in E_1$, $m = 1, 2, \cdots$). Since E_1 is a proper subsystem of E, the set $E_1^+ = E^+ \cap E_1$ contains an inner point y_0 of E_1 (by Lemma 2.2). Then $-y_0$ is an inner point of $E_1^- = E^- \cap E_1$.

Since E_1^- is evidently convex and x_0 belongs to the boundary of E_1^-, we have

$$\alpha(-y_0) + (1 - \alpha)x_0 \leqq 0,$$

whenever $0 < \alpha \leqq 1$, whence

$$\frac{1 - \alpha}{\alpha} x_0 \leqq y_0.$$

$(1 - \alpha)/\alpha$ takes on all positive values when α ranges through the semi-interval $(0, 1]$. We see that y_0 is an u.b. of Rx_0, x_0 being positive, contrary to the hypothesis that E is E_{H^*}. The existence of p_1 is proved. If $l(x)$ is an arbitrary linear function on E that coincides with p_1 on E_1 then $\varphi(x) = e^{il(x)}$ belongs to every $\Pi_{x_i} (i = 1, 2, \cdots, n)$.

We have proved that the intersection of an arbitrary finite system of closed sets Π_x $(x > 0)$ is not empty. Since Φ is bicompact, there exists $\varphi(x) = e^{ip(x)}$ that belongs to every Π_x $(x > 0)$.

It follows immediately from the definition of Π_x that $p(x)$ is a positive linear function coinciding with $p'(x)$ on $E'^+ = E^+ \cap E'$. Since E' is a proper subsystem of E, i.e. $E' = L(E'^+)$ (by Lemma 2.1), we obtain

$$p(x) = p'(x)$$

everywhere on E'. Theorem 2.4 is proved.

Applying Theorem 2.4 to the case $E' = D_{x_0}$, $p'(\lambda x_0) = \lambda$, where x_0 is an arbitrary positive element of a weakly homogeneous system E, we obtain

THEOREM 2.5. *If E is E_{H^*} then for every $x_0 > 0$ there exists a positive linear function $p(x)$ such that $p(x_0) = 1$.*

Let z be an arbitrary element of E incomparable with 0. According to Lemma 2.1, z can be represented in the form

$$z = x - y,$$

where $x > 0$, $y > 0$.[4] Consider the elements

$$z'(\alpha) = -\alpha x + y,$$
$$z''(\beta) = x - \beta y$$

and denote by σ (resp. τ) the least upper bound of the α (resp. β) such that $z'(\alpha) \geqq 0$ (resp. $z''(\beta) \geqq 0$). Since

$$z'(0) = y > 0, \qquad z''(0) = x > 0,$$
$$z'(1) = -x + y = -z \,||\, 0, \qquad z''(1) = x - y = z \,||\, 0,$$

we obtain the inequalities

$$0 \leqq \sigma \leqq 1,$$
$$0 \leqq \tau \leqq 1.$$

LEMMA 2.3. *Let E be E_{H^*} and $z \in E$ incomparable with 0. Then either*

$$\sigma = 1 \qquad (\tau = 1)$$

for every representation of z as the difference of two positive elements or

$$\sigma < 1 \ (\textit{resp. } \tau < 1)$$

for every representation of z.

[4] This representation is evidently not unique.

In the first case, R_z has an u.b. (resp. a l.b.) and for each positive linear function $p(x)$

$$p(z) \leqq 0 \ (\textit{resp. } p(z) \geqq 0).$$

In the second case, R_z has no u.b.'s (resp. l.b.'s) and there exists a positive linear function $p_0(x)$ such that

$$p_0(z) > 0 \ (\textit{resp. } p_0(z) < 0).$$

PROOF. 1. Suppose that

$$z = x - y \qquad (x > 0, y > 0)$$

and the corresponding $\sigma = 1$. It follows from the definition of σ that for every $\epsilon > 0$

$$-(1 - \epsilon)x + y \geqq 0,$$

whence

$$\epsilon x - x + y \geqq 0,$$

$$z = x - y \leqq \epsilon x,$$

$$\frac{1}{\epsilon} z \leqq x.$$

Since $\epsilon > 0$ was arbitrary, we obtain

$$R_z \leqq x.$$

Quite analogously $\tau = 1$ implies $R_z \geqq -y$.

2. If $R_z \leqq x_0$ ($R_z \geqq y_0$) and $p(x)$ is an arbitrary positive linear function then

$$p(\lambda z) = \lambda p(z) \leqq p(x_0) \qquad (\text{resp. } \geqq p(y_0))$$

for every $\lambda > 0$, whence

$$p(z) \leqq 0 \qquad (\text{resp. } p(z) \geqq 0).$$

3. Suppose that

$$z = x - y \qquad (x > 0, y > 0),$$

σ and τ being arbitrary. Take $E' = L(x, y)$ and define a linear function p' on E', putting

$$p'(x) = 1, \qquad p'(y) = \eta \geqq 0,$$

where η is not yet determined. It is easy to see that $p'(u)$ thus defined is a positive linear function on E' if and only if η satisfies the inequalities

$$\sigma \leqq \eta \leqq \frac{1}{\tau}. \tag{$\circ$}$$

We have, indeed, for every $\epsilon > 0$

$$z'(\sigma - \epsilon) = -(\sigma - \epsilon)x + y \geqq 0,$$
$$z''(\tau - \epsilon) = x - (\tau - \epsilon)y \geqq 0.$$

If p' is non-negative on $E'^{+} = E^{+} \cap E'$ then

$$p'(z'(\sigma - \epsilon)) = -\sigma + \epsilon + \eta \geqq 0,$$
$$p'(z''(\tau - \epsilon)) = 1 - (\tau - \epsilon)\eta \geqq 0.$$

Putting $\epsilon \rightarrow 0$, we obtain

$$-\sigma + \eta \geqq 0,$$
$$1 - \tau\eta \geqq 0,$$

or

$$\sigma \leqq \eta \leqq \frac{1}{\tau}.$$

The necessity of ($\circ$) is proved. It is not more difficult to prove the sufficiency of it. ($\circ$) includes also the case $\tau = 0$ if we put $1/0 = +\infty$.

Suppose that there exists a representation

$$z = x - y \qquad (x > 0, y > 0)$$

for which $\sigma < 1$. Then it is possible to choose η satisfying ($\circ$) and such that $\sigma \leqq \eta < 1$. By Theorem 2.4 p' can be extended to a positive linear function p on E. We have

$$p(z) = p'(z) = p'(x - y) = 1 - \eta > 0.$$

If for a certain representation of z

$$z = x - y \qquad (x > 0, y > 0)$$

$\tau < 1$, then $1/\tau > 1$ and we can choose $\eta > 1$ satisfying the inequalities ($\circ$). Then

$$p(z) = p'(z) = p'(x - y) = 1 - \eta < 0,$$

where p is positive linear function on E that coincides with p' on E'

Combining the results 1., 2. and 3., we obtain the statements of our Lemma.

§3. Archimedean systems and systems of real functions

Let E be E_A. The following lemmas state some characteristic properties of Archimedean systems.

LEMMA 3.1. *E is E_A if and only if for every $x > 0$ the null-element 0 is the g.l.b. of the set R_x.*

Proof. Suppose that E is E_A. For every $x > 0$

$$0 < R_x.$$

It suffices to show that

$$y \leqq R_x$$

implies

$$y \leqq 0.$$

If

$$y \leqq \lambda x$$

for arbitrary $\lambda > 0$ then

$$\mu y \leqq x,$$

where $\mu = 1/\lambda$ can take on all positive values. According to Axiom A

$$y \leqq 0.$$

Conversely, assume that $0 = \inf R_y$, whenever $y > 0$. If for a certain $x \in E$ there exists an u.b. of R_x then we may suppose that y is positive.

$$R_x \leqq y$$

implies obviously

$$x \leqq R_y$$

and by our hypothesis

$$x \leqq 0,$$

i.e. Axiom A is fulfilled.

Remark 3.1. It is almost evident that generally $0 = \inf_\alpha \{\lambda_\alpha x\}$, where $x > 0$, $\lambda_\alpha > 0$ and $\inf_\alpha \{\lambda_\alpha\} = 0$.

Lemma 3.2. *E is E_A if and only if $\lambda_n x \leqq y$, $\lambda_n \to \lambda$, imply $\lambda x \leqq y$.*

Proof. The sufficiency of the condition is almost evident. If

$$R_x \leqq y$$

then

$$R_y \geqq x$$

and

$$R_{-y} \leqq -x.$$

For $\lambda_n = 1/n$ we have then

$$\frac{1}{n}(-y) \leqq -x.$$

Taking $n \to \infty$ we obtain then

$$0 \leqq -x$$

or

$$x \leqq 0,$$

i.e. Axiom A is fulfilled.

Conversely, suppose that E is E_A and consider elements $x \in E$, $y \in E$ and a sequence $\{\lambda_n\}$ converging to λ such that

$$\lambda_n x \leqq y \qquad (n = 1, 2, \cdots).$$

Without loss of generality we can assume that $\{\lambda_n\}$ is a non-decreasing sequence, i.e.

$$\lambda_n = \lambda - \epsilon_n ,$$

where $\epsilon_1 \geqq \epsilon_2 \geqq \cdots$ and $\epsilon_n \to 0 (n \to \infty)$. If this is not the case, then we take a non-decreasing subsequence $\{\lambda_{n_k}\} \subset \{\lambda_n\}$ instead of $\{\lambda_n\}$; if $\{\lambda_n\}$ does not contain any such subsequences then we represent $\lambda_n x$ as $(-\lambda_n)(-x)$ and consider the sequence $\{-\lambda_n\}$.

We have

$$(\lambda - \epsilon_n)x \leqq y,$$

$$\lambda x - y \leqq \epsilon_n x.$$

Since $\epsilon_n \to 0 (n \to \infty)$,

$$0 = \inf \{\epsilon_n x; n = 1, 2, \cdots \}$$

and consequently

$$\lambda x - y \leqq 0,$$

$$\lambda x \leqq y,$$

as was to be proved.

LEMMA 3.3. *If E is an Archimedean system, then for every $z \,||\, 0$ and every representation of z as the difference of two positive elements*

$$z = x - y \qquad (x > 0, y > 0)$$

there is always $\sigma < 1$ and $\tau < 1$.

PROOF. Suppose that there exists an element $z \in E$ incomparable with 0 that can be so represented

$$z = x - y \qquad (x > 0, y > 0)$$

that, for instance, $\sigma = 1$. It follows from the definition of σ that for any $\alpha_n < 1$

$$-\alpha_n x + y \geqq 0 \qquad (n = 1, 2, \cdots).$$

Putting $\alpha_n \to 1 (n \to \infty)$ and applying Lemma 3.2, we obtain

$$-x + y \geqq 0$$

which is impossible, because $-x + y = -z \| 0$.

As to the inequality

$$\tau < 1$$

it can be proved in the same way.

Let $P = \{p\}$ be an arbitrary set, F an arbitrary linear system of real functions $f(p)$ defined on P. We define the *natural ordering* of F: $f_1 < f_2$ if $f_1(p) \leqq f_2(p)$ for all $p \in P$ and $f_1(p_0) < f_2(p_0)$ for at least one $p_0 \in P$. We assume that for every two functions $f_1 \in F, f_2 \in F$ there exists a function $f_3 \in F$ such that $f_3 \geqq f_1$, $f_3 \geqq f_2$.

THEOREM 3.1. *F is an Archimedean system. Conversely, every E_A is isomorphic to a system of real functions defined on some set P with the natural ordering.*

We say that the partially ordered linear system $E = \{x\}$ and $E_1 = \{y\}$ are *isomorphic* if there exists a one-to-one correspondence $y = \varphi(x)$ between the elements of E and E_1 such that $\varphi(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 \varphi(x_1) + \lambda_2 \varphi(x_2)$ and $x_1 < x_2$ implies $\varphi(x_1) < \varphi(x_2)$ and vice versa.

PROOF OF THEOREM 3.1. The first statement is almost trivial. Suppose that there exist $f_1 \in F$, $f_2 \in F$ such that

$$\lambda f_1 \leqq f_2$$

for all $\lambda > 0$. We then have

$$\lambda f_1(p) \leqq f_2(p)$$

for every $p \in P$. Since $\lambda > 0$ is arbitrary, this inequality implies

$$f_1(p) \leqq 0,$$

i.e. $f_1 \leqq 0$ in the sense of the natural ordering of F.

Let E be an arbitrary Archimedean system. Denote by P the totality of all positive linear functions on E. To every element $x \in E$ we let correspond the function $f_x(p) = p(x)$ defined on P.

The correspondence

$$x \to f_x(p)$$

is a homomorphism,

$$\alpha x + \beta y \to f_{\alpha x + \beta y}(p) = p(\alpha x + \beta y) = \alpha p(x) + \beta p(y) = \alpha f_x(p) + \beta f_y(p).$$

Furthermore it is biunivoque. It is sufficient to prove that $f_x \neq 0$, whenever $x \neq 0$, i.e. for every $x \neq 0$ there exists a positive linear function p_0 such that $f_x(p_0) = p_0(x) \neq 0$. For $x > 0$ or $x < 0$ it follows from Theorem 2.5. For $x \| 0$ it can be deduced from Lemmas 2.3 and 3.3.

Let $x > 0$. Then for every $p \in P$ $f_x(p) = p(x) \geqq 0$ and by Theorem 2.5 there exists a $p_0 \in P$ such that $f_x(p_0) = p_0(x) > 0$, i.e. $f_x > 0$.

We must show that for every $x \bar{\in} E^+$ $f_x \not> 0$, i.e. there exists such $p_1 \in P$ that $f_x(p_1) = p_1(x) < 0$. It follows from Theorem 2.5 in the case $x < 0$ and from Lemmas 2.3 and 3.3 if $x \parallel 0$.

Thus we see that $x > 0$ if and only if f_x is positive in the sense of natural ordering of $F = \{f_x\}$.

§4. The structure of weakly Archimedean systems

LEMMA 4.1. *If the weakly homogeneous system E is regular (i.e. is E_{A^*}) then for every $z \parallel 0$ and every representation*

$$z = x - y \qquad (x > 0, y > 0)$$

either σ and τ are equal to 1, or $\sigma < 1$ and $\tau < 1$ simultaneously.

PROOF. Assume $z \parallel 0$ and

$$z = x - y \qquad (x > 0, y > 0).$$

Suppose that $\sigma = 1$. According to Lemma 2.3 R_z has an u.b. Since E is regular, there exists also a l.b. of R_z. Applying the same Lemma 2.3, we conclude that $\tau = 1$.

$\tau = 1$ implies $\sigma = 1$ analogously. Lemma 4.1 is proved.

Let E be E_{A^*} and p an arbitrary positive linear function on E. Denote by W_p the set of $z \in E$ for which

$$p(z) = 0$$

and by W—the intersection of all W_p,

$$W = \bigcap_{p}^{P} W_p.$$

W is evidently a linear subsystem of E and every element of W is incomparable with 0.

It follows from Lemmas 2.3 and 4.1 that W consists of the elements $z \in E$, $z \parallel 0$, for which D_z is bounded and for every representation

$$z = x - y \qquad (x > 0, y > 0)$$

$\sigma = \tau = 1$.

As to the elements $z \parallel 0$ that do not belong to W, for every such z and each representation

$$z = x - y \qquad (x > 0, y > 0)$$

$\sigma < 1$ and $\tau < 1$. By Lemma 2.3 for a given $z \parallel 0$, $z \bar{\in} W$, there exist such $p_1 \in P$, $p_2 \in P$ that $p_1(z) > 0$, $p_2(z) < 0$.

LEMMA 4.2. *If E is E_{A^*,C^*} then E is an Archimedean system.*

PROOF. Suppose that E satisfies Axiom A^* and is a lattice. If x and y are elements of E such that

$$R_x \leqq y$$

then according to Axiom A^* we must have either

$$x \leqq 0$$

or

$$x \,||\, 0.$$

In the second case we would have

$$\lambda(y \vee 0) = (\lambda y) \vee 0 \leqq x \vee 0$$

(see [1]), where $y \vee 0 > 0$, for all $\lambda > 0$ which contradicts Axiom A^*. Lemma 4.2 is proved.

Let $R = \{r\}$ be an arbitrary set and G a linear system of real functions $g(r)$ defined on R. We introduce a *generalized natural ordering* of G: a subset $P \subset R$ is fixed and $g_1 < g_2$ denotes that $g_1(p) \leqq g_2(p)$ for all $p \in P$ and $g_1(p_0) < g_2(p_0)$ for at least one $p_0 \in P$.

We call P the *fundamental subset* of R.

We assume that G satisfies Axiom III.

THEOREM 4.1. *G satisfies Axiom A^*. G is G_A if and only if $g = 0$ is the unique function that vanishes identically on the fundamental subset $P \subset R$.*

Conversely, every E_{A^*} is isomorphic to a linear system of real functions $G = \{g\}$ defined on some set R with a generalized natural ordering. The functions $g \in G$ that correspond to the elements of the subsystem $W \subset E$ are those that are equal to 0 identically on the fundamental subset $P \subset R$.

PROOF. If $g_1 \in G$, $g_2 \in G$ satisfy the inequality

$$\lambda g_1 \leqq g_2$$

for any positive λ then necessarily

$$g_1(p) \leqq 0,$$

whenever $p \in P$. Consequently, the following cases are possible:

1) $g_1 = 0$ if $g_1(r) \equiv 0$ $(r \in R)$;

2) $g_1 < 0$ if there exists a $p_0 \in P$ such that $g_1(p_0) < 0$;

3) $g_1 \,||\, 0$ if $g_1(p) \equiv 0 (p \in P)$ and $g_1(r_0) \neq 0$ for at least one $r_0 \in R \dot{-} P$. We see that Axiom A^* is fulfilled.

G satisfies Axiom A if and only if the third case ($g_1 \,||\, 0$) is excluded, for which the following condition is necessary and sufficient: there exist no $g \in G$ such that $g \neq 0$ and $g(p) \equiv 0$ $(p \in P)$.

Let E be an arbitrary system satisfying Axiom A^*. Suppose that

$$B = \{x_0, x_1, \cdots, x_\nu, \cdots\}, \qquad x_\nu \in E, \qquad \nu < \vartheta,$$

is a transfinite sequence of elements such that

1) $E = L(B)$,
2) for every ordinal $\nu < \vartheta$: $x_\nu \bar{\in} L(\{x_\mu\})$, where $0 \leqq \mu < \nu$,
3) there exists an ordinal $\vartheta' < \vartheta$ such that $W = L(\{x_\nu\})$, where $0 \leqq \nu < \vartheta'$.

Denote by P the set of all positive linear functions p on E and by Q the set of such linear functions q on E that

$$q(x_\nu) = 0,$$

whenever $\vartheta' \leqq \nu < \vartheta$. It follows from the definition of W that $P \cap Q$ is empty.

We introduce the notation $\boldsymbol{R}$ for $P \cup Q$. To every $x \in E$ we let correspond the function $g_x(r)$ defined on $R = \{r\}$,

$$x \to g_x(r) = r(x).$$

We have evidently

$$\alpha x + \beta y \to g_{\alpha x + \beta y}(r) = r(\alpha x + \beta y) = \alpha r(x) + \beta r(y) = \alpha g_x(r) + \beta g_y(r).$$

Introduce the generalized natural ordering of G determined by the fundamental subset $P \subset R$. We shall prove that $x > 0$ is equivalent to $g_x > 0$. It follows from Theorem 2.5 that $x > 0$ ($x < 0$) implies $g_x > 0$ (resp. $g_x < 0$). Suppose that $x \,||\, 0$. If $x \bar{\in} W$ then there exist such $p_1 \in P$, $p_2 \in P$ that $g_x(p_1) = p_1(x) > 0$, $g_x(p_2) = p_2(x) < 0$. If $x \in W$ then $g_x(p) \equiv p(x) \equiv 0$ ($p \in P$), but for at least one $q_0 \in Q$: $g_x(q_0) = g_0(x) \neq 0$. In both cases $g_x \,||\, 0$.

Incidentally, we have proved that the correspondence $x \to g_x(r)$ is biunivoque. Theorem 4.1 is proved.

Let $\tilde{E}$ be the factor system $E \ominus W$—the system of classes $\tilde{x} = x + W$, where $x \in E$. $\tilde{E}$ is a linear system. We define an ordering of $\tilde{E}$ as follows:

$\tilde{x}_1 < \tilde{x}_2$ if there exist such $x_1 \in \tilde{x}_1$, $x_2 \in \tilde{x}_2$ that $x_1 < x_2$.

Applying Theorem 4.1 it is easy to see that this is an actual ordering of $\tilde{E}$, i.e. the relation of order defined between the classes $\tilde{x}_1$, $\tilde{x}_2$ does not depend on the choice of the representatives x_1, x_2.

THEOREM 4.2. *$\tilde{E}$ is an Archimedean system.*

THEOREM 4.3. *If E is the direct sum of W and a subsystem $E_1 \subset E$, i.e. for every $x \in E$ there exists the unique representation*

$$x = z + y, \qquad z \in W, \qquad y \in E_1,^5$$

then E_1 is an Archimedean system and the ordering of E can be re-defined as follows: $x_1 < x_2$ if for $x_1 = z_1 + y_1$, $x_2 = z_2 + y_2$ $y_1 < y_2$.

One can easily prove these theorems applying Theorem 4.1.

---

[5] This is not the case in general.

§5. Closed systems

A very important class of partially ordered linear systems is that of closed systems.

Axiom C guarantees the existence of the l.u.b. for every set $M \subset E$ that has an u.b. Suppose that a set $M \subset E$ has a l.b. Then the set $N(M)$ is not empty and has u.b.'s (each $x \in M$ is one of them). Consequently there exists sup $N(M)$ which is evidently the g.l.b. of M. We see that the set M has inf M if it has a l.b.

The following lemma will be useful.

LEMMA 5.1. *Let E be E_c. For any two subsets $M_1 \subset E$, $M_2 \subset E$ that have u.b.'s and $\lambda > 0$*

$$\sup (M_1 + M_2) = \sup M_1 + \sup M_2,$$
$$\sup (\lambda M_1) = \lambda \sup M_1.$$[6]

PROOF. Since

$$M_1 + M_2 \leqq \sup M_1 + \sup M_2,$$

we have

$$\sup (M_1 + M_2) \leqq \sup M_1 + \sup M_2.$$

On the other hand

$$M_1 + M_2 \leqq \sup (M_1 + M_2)$$

implies

$$M_1 \leqq \sup (M_1 + M_2) - M_2,$$
$$\sup M_1 \leqq \sup (M_1 + M_2) - M_2,$$
$$M_2 \leqq \sup (M_1 + M_2) + \sup M_1,$$
$$\sup M_2 \leqq \sup (M_1 + M_2) - \sup M_1$$

and

$$\sup M_1 + \sup M_2 \leqq \sup (M_1 + M_2).$$

The proof of the second equality is not more difficult.

THEOREM 5.1. *Every closed system is Archimedean.*

PROOF. Suppose that E is E_c and E contains such elements x and y that

$$R_x \leqq y.$$

If $z = \sup R_x$ then by Lemma 5.1 for an arbitrary $\lambda > 0$

$$\lambda z = \sup (\lambda R_x).$$

---

[6] If E is not closed then these equalities hold, whenever the right sides exist.

Since $\lambda R_x = R_x$, we have

$$\lambda z = z,$$

whence

$$z = 0.$$

Consequently

$$R_x \leqq 0$$

and in particular

$$x \leqq 0,$$

as was to be proved.

Let E be an arbitrary partially ordered linear system. We say that the partially ordered linear system E_1 is the *extension* of E if E is isomorphic to a subsystem $E' \cap E_1$. If E_1 is closed then we say that E_1 is a *closed extension* of E.

The problem of the existence of closed extensions of partially ordered linear systems has been studied by A. Youdine [6]. The following theorem contains the solution of the problem from another point of view.

THEOREM 5.2. *The necessary and sufficient condition for the existence of closed extensions of a partially ordered linear system E is that E be Archimedean. If E is E_A then among all closed extensions E_1 of E there exists a minimal one $\hat{E}$ such that every other E_1 is the extension of $\hat{E}$.*

PROOF. The necessity of the condition follows from Theorem 5.1 and the cogradiency of Axiom A.

In order to prove the sufficiency of the condition we shall use the results obtained by H. M. MacNeille who extended the Dedekind's method to arbitrary partially ordered sets [F].

Suppose that E is E_A and let $\hat{E}$ be the totality of all "cuts" (S, T) in E. The cut (S, T) is the pair of subsets $S \subset E$, $T \subset E$ such that

1) S, T are not empty,

2) $S = N(T)$ and $T = H(S)$.

The ordering of $\hat{E}$ is defined as follows:

$$(S, T) \leqq (S', T') \text{ if } S \subset S' \text{ (consequently } T \supset T').$$

The partially ordered set $\hat{E}$—we call it *closure* of E—possesses the following properties:

1. $\hat{E}$ is closed, i.e. it satisfies Axiom C.
2. The correspondence

$$x \rightarrow (S_x, T_x),$$

where $S_x = N(x)$, $T_x = H(x)$, is an isomorphism[7] between $E = \{x\}$ and the subset $E' \subset \hat{E}$ consisting of cuts of such kind.

[7] Since algebraic operations are not yet defined in $\hat{E}$, we understand here by *isomorphism* simply an ordering-preserving one-to-one correspondence.

3. If a subset $M \subset E$ has the l.u.b. (g.l.b.) $a \in E$ then the l.u.b. (resp. g.l.b.) of the set $M' \subset \hat{E}$ corresponding to M and considered as the subset of $\hat{E}$ is the cut (S_a, T_a) corresponding to a.

4. Every closed partially ordered set E_1 that contains a subset E_2 isomorphic to E contains the subset $\hat{E}_2 \supset E_2$ isomorphic to $\hat{E}$. If

$$x \to \varphi(x) \in E_2,$$

$$\xi \to \psi(\xi) \in \hat{E}_2,$$

where $x \in E$, $\xi \in \hat{E}$, are these isomorphisms, then for $\xi = (S_x, T_x)$

$$\varphi(x) = \psi(\xi).$$

We shall consider $\hat{E}$ as consisting of some elements ξ set into a one-to-one correspondence with the cuts (S, T) which will be denoted by

$$\xi \sim (S, T)$$

or

$$\xi \sim S,$$

where S is the lower class of the cut (S, T). The elements ξ corresponding to the "rational" cuts (S_x, T_x) will be identified with the elements $x \in E$. $\xi \leqq \xi'$ denotes that $S \subset S'$ if $\xi \sim S$, $\xi' \sim S'$.

Denote by $\mathfrak{S}$ and $\mathfrak{T}$ the systems of all sets S and T respectively. We shall use the following properties of the sets $S \in \mathfrak{S}$ (to everyone of them corresponds the dual property of the sets $T \in \mathfrak{T}$). The properties 1^0–4^0 hold for every partially ordered set E, the other ones deal with algebraic operations in E.

1^0. The sets $S \in \mathfrak{S}$ can be characterized by the following properties: S is not empty, has an u.b. and

$$S = N(H(S)).^8$$

2^0. If $x \in S$ and $x' < x$ then $x' \in S$.

3^0. If $\{S_\alpha\}$ is an arbitrary subsystem of $\mathfrak{S}$ and the intersection $S_0 = \bigcap_\alpha S_\alpha$ is not empty then $S_0 \in \mathfrak{S}$.

4^0. For every non-empty set $M \subset E$ that has an u.b. there exists the minimal set of the system $\mathfrak{S}$—denote it by $[M]$—that contains M and $[M] = N(H(M))$.

5^0. If $x \in E$ and $\lambda > 0$ then for every $S \in \mathfrak{S}$ the sets $S + x$ and λS belong to $\mathfrak{S}$.

6^0. If $S \in \mathfrak{S}$, $T \in \mathfrak{T}$ and $\lambda < 0$ then $\lambda S \in \mathfrak{T}$ and $\lambda T \in \mathfrak{S}$.

7^0. If the subset $M \subset E$ has an u.b. then $[M + y] = [\mathfrak{M}] + y$ and $[\lambda M] = \lambda[M]$, whenever $y \in E$ and $\lambda > 0$.

8^0. Every $S \in \mathfrak{S}$ is a convex set.

The property 1^0 can be easily deduced from the definition of the cut.

2^0 is evident.

---

[8] It is obvious that for every $M \subset E$ $M \subset N(H(M))$.

$\mathfrak{S}$ is isomorphic to $\hat{E}$ (if $S_1 \leqq S_2$ denotes that $S_1 \subset S_2$) and the closedness of $\hat{E}$ is based upon 3^0 which therefore is worth proving. Suppose that S_0 is not empty. Then it has an u.b., since $S_0 \subset S_\alpha$ and S_α has one. For every S_α

$$S_0 \subset S_\alpha = N(H(S_\alpha))$$

and consequently

$$H(S_0) \supset H(S_\alpha),$$

whence

$$N(H(S_0)) \subset N(H(S_\alpha)) = S_\alpha .$$

Since S_α was arbitrary, we have

$$N(H(S_0)) \subset S_0 .$$

On the other hand it is evident that

$$N(H(S_0)) \supset S_0 .$$

According to 1^0 $S_0 \in \mathfrak{S}$.

The first part of 4^0 follows immediately from 3^0: $[M]$ is the intersection of all S that contain M. Let us prove that $[M] = N(H(M))$. Consider the sets $[M] = \bigcap_{S \supset M} S$ and $S' = N(H(M))$. They both belong to $\mathfrak{S}$. Since

$$M \subset S',$$

we have

$$[M] \subset S'.$$

On the other hand

$$M \subset [M],$$

whence

$$H(M) \supset H([M])$$

and

$$S' = N(H(M)) \subset N(H([M])) = [M].$$

Consequently

$$[M] = N(H(M)).$$

5^0 and 6^0 can be easily proved with the use of 1^0.

If M has an u.b. and $y \in E$ then

$$[M] + y \supset M + y.$$

Since $[M] + y$ belongs to $\mathfrak{S}$ by 5^0, we have

$$[M] + y \supset [M + y].$$

Suppose that there exists such $x \in [M] + y$ that $x \bar{\in} [M + y]$. Since

$$x = a + y,$$

where $a \in [M]$, we would have

$$a \bar{\in} [M + y] - y.$$

The set $[M + y] - y$ belongs to $\mathfrak{S}$ and contains M. Consequently

$$[M + y] - y \supset [M] \supset a$$

and that contradicts our hypothesis. The second part of 7^0 can be proved in a similar way.

We now prove 8^0. Let x_1, x_2 be arbitrary elements of the set S. Then

$$x_1 \leqq T = H(S),$$

$$x_2 \leqq T = H(S),$$

i.e.

$$x_1 \leqq y,$$

$$x_2 \leqq y,$$

whenever $y \in T$. If λ and μ are non-negative numbers with $\lambda + \mu = 1$ then

$$\lambda x_1 \leqq \lambda y,$$

$$\mu x_2 \leqq \mu y,$$

and

$$\lambda x_1 + \mu x_2 \leqq \lambda y + \mu y = (\lambda + \mu) y = y.$$

Since $y \in T$ was arbitrary,

$$\lambda x_1 + \mu x_2 \leqq T$$

or

$$\lambda x_1 + \mu x_2 \in N(T) = S,$$

as was to be proved.

Lemma 5.1. *For every cut (S, T) in E there exists inf $(T - S)$ which is equal to 0 if and only if E is E_A.*

Proof. Without restricting generality we can suppose that $0 \in A$. If it were not so, we could "translate" the cut taking $(S - x_0, T - x_0)$ instead of (S, T), where $x_0 \in S$.

It is evident that $T - S \geqq 0$. Suppose that $T - S \geqq z$. In order to prove that 0 is the g.l.b. of $T - S$, we must show that $z \leqq 0$.

$$T - S \geqq z$$

implies

$$S + z \leqq T$$

or

$$S + z \subset S.$$

It is easy to see that

$$S + nz \subset S$$

for every integer n. Since $0 \in S$, we have in particular

$$0 + \{nz\} \subset S,$$
$$\{nz\} \subset S,$$
$$\{nz\} \leqq T.$$

It follows from Remark 1.2 that

$$R_z \leqq T,$$

whence by Axiom A

$$z \leqq 0.$$

If E is not E_A, i.e. E contains an element z positive or incomparable with 0 such that R_z has an u.b., then it is not difficult to see that the cut (S, T), where $S = [R_z]$, does not possess the property stated in Lemma 5.1.

After these preliminaries we can prove Theorem 5.2. We shall show that it is possible to define the algebraic operations in $\hat{E}$ in such way that $\hat{E}$ converts into a partially ordered linear system containing E with the a priori given operations as its subsystem.

Definition O. *If* $\xi_1 \sim S_1$, $\xi_2 \sim S_2$ *then we put*

$$\xi_1 + \xi_2 \sim [S_1 + S_2].$$

If $\xi \sim (S, T)$ *then*

$$\lambda\xi \sim (\lambda S, \lambda T) \qquad \text{if } \lambda > 0,$$
$$\lambda\xi \sim (N(0), H(0)) = 0 \qquad \text{if } \lambda = 0,$$
$$\lambda\xi \sim (\lambda T, \lambda S) \qquad \text{if } \lambda < 0.$$

$\hat{E}$ with the operations just defined is a linear system. Let

$$\xi \sim S_1, \qquad \eta \sim S_2, \qquad \zeta \sim S_3$$

be arbitrary elements of $\hat{E}$.

I. $(\xi + \eta) + \zeta = \xi + (\eta + \zeta)$.

According to Definition O

$$(\xi + \eta) + \zeta \sim [[S_1 + S_2] + S_3],$$
$$\xi + (\eta + \zeta) \sim [S_1 + [S_2 + S_3]].$$

Since

$$S_1 + S_2 \subset [S_1 + S_2],$$

we have

$$S_1 + S_2 + S_3 \subset [S_1 + S_2] + S_3\,,$$

whence

$$[S_1 + S_2 + S_3] \subset [[S_1 + S_2] + S_3].$$

On the other hand, if

$$x \in H(S_1 + S_2 + S_3),$$

i.e.

$$x \geqq S_1 + S_2 + S_3\,,$$

then

$$x - S_3 \geqq S_1 + S_2$$

and by 4^0

$$x - S_3 \geqq [S_1 + S_2],$$

$$x \geqq [S_1 + S_2] + S_3$$

$$x \geqq [[S_1 + S_2] + S_3].$$

We see that

$$H(S_1 + S_2 + S_3) \subset H([S_1 + S_2] + S_3]),$$

whence

$$[S_1 + S_2 + S_3] = N(H(S_1 + S_2 + S_3))$$
$$\supset N(H([[S_1 + S_2] + S_3])) = [[S_1 + S_2] + S_3].$$

Consequently

$$[[S_1 + S_2] + S_3] = [S_1 + S_2 + S_3].$$

Analogously we can obtain

$$[S_1 + [S_2 + S_3]] = [S_1 + S_2 + S_3],$$

whence

$$[[S_1 + S_2] + S_3] = [S_1 + [S_2 + S_3]]$$

or

$$(\xi + \eta) + \zeta = \xi + (\eta + \zeta).$$

II. $\xi + \eta = \eta + \xi$.
We have, indeed, $S_1 + S_2 = S_2 + S_1$, whence $[S_1 + S_2] = [S_2 + S_1]$.
III. $0 \sim N(0)$ is the null-element of $\hat{E}$, i.e. $\xi + 0 = \xi$ for every $\xi \in \hat{E}$ or $[S + N(0)] = S$.
If $x \in S$, $y \in N(0)$ then $x + y \leqq x$. According to 2^0 $x + y \in S$. We see that

$$S + N(0) \subset S,$$

whence

$$[S + N(0)] \subset S.$$

Conversely, since $0 \in N(0)$, we have

$$S = S + 0 \subset S + N(0) \subset [S + N(0)].$$

IV. For every $\xi \in \hat{E}$ there exists such $-\xi \in \hat{E}$ that $\xi + -\xi = 0$.
Let $\xi \sim (S, T)$. We shall show that

$$-\xi = (-1)\xi \sim (-T, -S).$$

According to Definition O

$$\xi + -\xi \sim [S + (-T)] = [S - T].$$

Since $S - T \leqq 0$ or $S - T \subset N(0)$, we have

$$[S - T] \subset N(0).$$

Lemma 5.1 gives

$$\inf (T - S) = 0,$$

whence

$$\sup (S - T) = 0.$$

We obtain $x \geqq 0$, whenever $x \geqq S - T$, i.e.

$$H(S - T) \subset H(0),$$

whence by 1^0 and 4^0

$$[S - T] = N(H(S - T)) \supset N(N(0)) = N(0).$$

We obtain

$$[S - T] = N(0)$$

or

$$\xi + -\xi = 0.$$

V. $1 \cdot \xi = \xi$,
that is obvious.
VI. $\lambda(\mu\xi) = (\lambda\mu)\xi$ for arbitrary real numbers λ, μ and $\xi \sim (S, T)$.

If either λ or μ is equal to 0 then the equality is obvious. Consider, for instance, the case: $\lambda < 0$, $\mu > 0$. Then $\lambda\mu < 0$ and by Definition O

$$\mu\xi \sim (\mu S, \mu T),$$

$$\lambda(\mu\xi) \sim (\lambda(\mu T), \lambda(\mu S)),$$

$$(\lambda\mu)\xi \sim ((\lambda\mu)T, (\lambda\mu)S).$$

We see that $\lambda(\mu\xi)$ and $(\lambda\mu)\xi$ correspond to the same cut, i.e. $\lambda(\mu\xi) = (\lambda\mu)\xi$. One can consider the other cases analogously.

VII. $(\lambda + \mu)\xi = \lambda\xi + \mu\xi$, where $\xi \sim S$ and λ, μ are arbitrary.

If one of the numbers λ, μ is equal to 0, then the equality is obvious. Consider the case: $\lambda > 0$, $\mu > 0$. According to Definition O

$$(\lambda + \mu)\xi \sim (\lambda + \mu)S,$$

$$\lambda\xi + \mu\xi \sim [\lambda S + \mu S].$$

It is evident that

$$(\lambda + \mu)S \subset \lambda S + \mu S.$$

Conversely, if $x \in \lambda S + \mu S$, i.e.

$$x = \lambda x_1 + \mu x_2,$$

where $x_1 \in S$, $x_2 \in S$, then

$$x = (\lambda + \mu)\left(\frac{\lambda}{\lambda + \mu} x_1 + \frac{\mu}{\lambda + \mu} x^2\right).$$

According to 8^0

$$x_3 = \frac{\lambda}{\lambda + \mu} x_1 + \frac{\mu}{\lambda + \mu} x_2 \in S$$

and

$$x = (\lambda + \mu)x_3 \in (\lambda + \mu)S.$$

We obtain

$$\lambda S + \mu S \subset (\lambda + \mu)S,$$

consequently

$$[\lambda S + \mu S] = \lambda S + \mu S = (\lambda + \mu)S.$$

Using I, II, IV and VI we can reduce the other cases to the one just considered.

VIII. $\lambda(\xi + \eta) = \lambda\xi + \lambda\eta$.

Let $\xi \sim (S_1, T_1)$, $\eta \sim (S_2, T_2)$. The case: $\lambda = 0$, is trivial. Suppose that $\lambda > 0$. We have

$$\lambda(\xi + \eta) \sim \lambda[S_1 + S_2],$$

$$\lambda\xi + \lambda\eta \sim [\lambda S_1 + \lambda S_2].$$

It is evident that

$$\lambda S_1 + \lambda S_2 = \lambda(S_1 + S_2),$$

whence by 5^0

$$[\lambda S_1 + \lambda S_2] = [\lambda(S_1 + S_2)] = \lambda[S_1 + S_2].$$

The case: $\lambda = -1$, follows essentially from IV. The case: $\lambda < 0$, can be considered as the combination of the previous ones.

Suppose now that

$$\xi \sim S_1, \qquad \eta \sim S_2, \qquad \zeta \sim S_3$$

and $\xi \leqq \eta$, i.e. $S_1 \subset S_2$. Then

$$S_1 + S_3 \subset S_2 + S_3,$$

whence

$$[S_1 + S_3] \subset [S_2 + S_3]$$

or

$$\xi + \zeta \leqq \eta + \zeta.$$

If $\lambda > 0$ then

$$S_1 \subset S_2$$

implies

$$\lambda S_1 \subset \lambda S_2.$$

Since $\lambda\xi \sim \lambda S_1$, $\lambda\eta \sim \lambda S_2$, we obtain

$$\lambda\xi \leqq \lambda\eta.$$

We have proved that $\hat{E}$ satisfies Axioms I and II. Axiom III is also fulfilled, because $\hat{E}$ being closed is a lattice and for any two $\xi \in \hat{E}$, $\eta \in \hat{E}$ there exists $\eta = \xi \vee \eta$.

The operations in $\hat{E}$ introduced by Definition O being applied to the elements of E (we identify E with the set of "rational" cuts in E) coincide with the *a priori* given operations in E. This statement follows immediately from the evident equalities

$$N(x) + N(y) = N(x + y),$$

$$N(\lambda x) = \begin{cases} \lambda N(x), & \lambda > 0, \\ N(0), & \lambda = 0; \\ \lambda H(x), & \lambda < 0, \end{cases}$$

when x, y are arbitrary elements of E.

Suppose that the algebraic operations are defined in E so that E becomes a subsystem of $\hat{E}$. It follows from Lemma 5.1 that the operations in $\hat{E}$ necessarily coincide with those introduced in Definition O. This proves Theorem 5.2.

REMARK 5.1. The main result and the methods of this paragraph slightly modified hold for partially ordered groups. Axiom A for the multiplicative group $G = \{a\}$ can be formulated as follows:

If for an element $a \in G$ the set $\{a^n; n = 1, 2, \cdots\}$ has an u.b. then $a \leqq e$, where e is the unit.

§6. Systems E_{B^*}

Let E be E_{B^*}. The following lemma states some characteristic properties of axial elements.

LEMMA 6.1. *Each of the following conditions is necessary and sufficient in order that $u \in E$ be an axial element of E:*

(α) *For every $x \in E$ the inequalities*

$$-\lambda u \leqq x \leqq \lambda u$$

hold, whenever $\lambda > 0$ is sufficiently large [8].

(β) *For every $x \in E$ the inequalities*

$$-u \leqq \mu x \leqq u$$

hold, whenever $\mu > 0$ is sufficiently small.

PROOF. If u is an axial element of E then for every $x \in E$

$$x = \lambda' u - x',$$
$$-x = \lambda'' u - x'',$$

where $\lambda' \geqq 0$, $\lambda'' \geqq 0$, $x' \in E^+$, $x'' \in E^+$. Taking $\lambda \geqq \max\{\lambda', \lambda''\}$, we have

$$-\lambda u \leqq x \leqq \lambda u.$$

The converse is obvious.

(α) and (β) are evidently equivalent.

REMARK 6.1. Every E_{B^*} contains an infinity of axial elements. It is evident that if u is an axial element of E and $v > u$ then v is also an axial element.

Lemma 6.1 gives

LEMMA 6.2. *Let u be an axial element of E and $v > 0$. Then v is also an axial element of E if and only if*

$$\lambda u \leqq v,$$

whenever $\lambda > 0$ is sufficiently small.

THEOREM 6.1. *E is E_{B^*} if and only if E^+ contains a bounded set V such tha every $x \in E^+$ can be represented in the form $x = \lambda y$, where $\lambda \geqq 0$, $y \in V$.*

PROOF. If E is E_{B^*} then we put

$$V = E^+ \cap N(u),$$

where u is an axial element of E.

Conversely, if E^+ contains such V then every $u \geqq V$ is an axial element.

THEOREM 6.2. *A weakly Archimedean system E is E_{B^*} if and only if the factor system $\tilde{E} = E \ominus W$ is $\tilde{E}_{B^*}$.*

The proof follows immediately from the definition of $\tilde{E}$ (§4).

THEOREM 6.3. *An Archmidean system E is E_{B^*} if and only if its closure $\hat{E}$ is $\hat{E}_{B^*}$. If it is so, then the axial element of $\hat{E}$ can be chosen in E.*

It is easy to verify these statements using Remark 6.1.

THEOREM 6.4. *If E is E_{B^*,C^*} then every axial element $u \in E$ is a unit in the sense of H. Freudenthal, i.e. $u \wedge x > 0$, whenever $x > 0$.*

PROOF. Assume the contrary. We would have an element $x > 0$ such that

$$u \wedge x = 0.$$

Then for every $\lambda > 0$

$$(\lambda u) \wedge x = 0$$

and

$$\lambda u + x = (\lambda u) \vee x \qquad [1]$$

Owing to Lemma 6.1 $\lambda > 0$ can be chosen such that

$$\lambda u \geqq x.$$

Therefore

$$(\lambda u) \vee x = \lambda u$$

and we obtain

$$\lambda u + x = \lambda u,$$

$$x = 0,$$

which contradicts our hypothesis.

REMARK 6.2. The converse is not true: there exist lattices that have units and do not contain axial elements at all.

Let u be a fixed axial element, $y \in E$ and α an arbitrary positive number. Consider the set $U_\alpha(y)$ consisting of the element x with

$$-(\alpha - \epsilon)u \leqq y - x \leqq (\alpha - \epsilon)u,$$

where $\epsilon > 0$ depends on x. Every such set we call *interval* in E. Introduce the notation $\mathfrak{A}$ for the totality of all intervals in E.

THEOREM 6.5. *A topology can be defined in E satisfying the conditions:*

1. *E converts into a T_1-space* [9];
2. *algebraic operations in E are continuous;*
3. *$\mathfrak{A}$ is the basis of the space E;*

if and only if E is homogeneous. If E is E_R then the needed topology can be generated by a norm.

PROOF. Suppose that E is not homogeneous, i.e. E contains an element $z \neq 0$ with

$$x_1 \leqq D_z \leqq x_2 .$$

Then for every $\mu > 0$

$$\mu^2 x_1 \leqq \mu^2 D_z \leqq \mu^2 x_2 .$$

Since $\mu^2 D_z = D_z$,

$$\mu^2 x_1 \leqq D_z \leqq \mu^2 x_2 .$$

Putting μ small enough in order that

$$-u \leqq \mu x_i \leqq u \qquad (i = 1, 2),$$

where u is a fixed axial element of E, we obtain

$$-\mu u \leqq \mu^2 x_i \leqq \mu u$$

and

$$-\mu u \leqq D_z \leqq \mu u .$$

Since in these inequalities we can take μ as small as we please, D_z and z in particular belong to every $U_\mu(0)$. Therefore 1. cannot be fulfilled.

If E is E_H then we define for every $x \epsilon E$

$$\| x \|_u = \sup \{\mu; -u \leqq \mu x \leqq u\} \qquad [8].$$

For every $x \epsilon E$ $\| x \|_u$ is finite. It is an immediate application of Axiom H. Lemma 6.1 gives that $\| x \|_u = 0$ only if $x = 0$. Furthermore it is evidently

$$\| \lambda x \|_u = | \lambda | \, \| x \|_u ,$$

$$\| x + y \|_u \leqq \| x \|_u + \| y \|_u .$$

We see that $\| x \|_u$ possesses the properties of the norm, we call it *u-norm*. It is evident that $U_\alpha(y)$ is the (open) α-sphere with the center y in the sense of u-norm. This completes the proof of the theorem.

In the following we shall understand by "u-completeness", "u-convergence" etc. the corresponding notions based on the u-norm. It is evident that if u and v are different axial elements of E then u-norm and v-norm are topologically equivalent.

REMARK 6.3. If E is a u-normed lattice then

$$\| x \|_u = \sup \{\mu > 0; \mu \, | x | \leqq u\} = \inf \{\lambda > 0; | x | \leqq \lambda u\},$$

where

$$| x | = x \vee (-x) \qquad [2].$$

It is easy to deduce from this remark the following

LEMMA 6.3. *If E is u-normed and closed, x is the u-limit of a sequence $\{x_n\} \subset E$ then $\{x_n\}$ converges to x in the sense of L. Kantorovitch* [2].

THEOREM 6.6. *If E is u-normed and closed then E is u-complete.*

PROOF. Suppose that $\{x_n\}$ is such that for every $\epsilon > 0$

$$|x_m - x_n| \leqq \epsilon u,$$

whenever m and n are sufficiently large. According to [2], Theorem 20, $\{x_n\}$ has the limit x in the sense of L. Kantorovitch and the inequalities

$$-\epsilon u \leqq x_m - x_n \leqq \epsilon u$$

imply

$$-\epsilon u \leqq x - x_n \leqq \epsilon u,$$

that means x is the u-limit of $\{x_n\}$.

§7. Partially ordered Banach spaces

In this paragraph we consider partially ordered Banach spaces, actually those the unit sphere of which has an u.b. and a l.b.[9]

We assume that E is E_B.

LEMMA 7.1. *Every E_B is E_{B^*} and the u.b. of the unit sphere is an axial element of E_B.*

PROOF. E_B satisfies the condition of Theorem 6.1 if we put V equal to the intersection of E_B^+ with the unit sphere. Every u.b. of the unit sphere is an u.b. of V and consequently is an axial element of E_B.

The *a priori* given norm in E we call N-norm. The topological and metrical properties of E based on the N-norm such as "closure", "boundedness" etc. are called "N-closure", "N-boundedness" etc.

LEMMA 7.2. *Axiom B is equivalent to the following condition: E is a Banach space ordered in such a way that E^+ contains an N-inner point.*

PROOF. Denote by $S(y; \alpha)$ the (N-open) α-sphere with the center y. If E is E_B, i.e.

$$S(0; 1) \leqq y$$

then

$$y - S(0; 1) = y + S(0; 1) \geqq 0.$$

[9] Such is, for instance, the space C of continuous functions $x(t)$, $0 \leqq t \leqq 1$, with the natural ordering and the norm $\|x\| = max\,|x(t)|$. On the contrary, the space $L_p (p \geqq 1)$ of measurable functions $x(t)$, $0 \leqq t \leqq 1$, with summable $|x(t)|^p$ does not satisfy Axiom B if $\|x\| = \left(\int_0^1 |x(t)|^p\,dt\right)^{\frac{1}{p}}$ and $x < y$ denotes that $x(t) \leqq y(t)$ almost everywhere on $[0, 1]$ and $x(t) < y(t)$ on a set $E \subset [0, 1]$ of positive measure.

Since $y + S(0; 1) = S(y; 1)$, we obtain

$$S(y; 1) \geqq 0,$$

y is an N-inner point of E^+.

Conversely, if y is an N-inner point of E^+, i.e.

$$S(y; \epsilon) \geqq 0,$$

then

$$S(y; \epsilon) = y + S(0; \epsilon) = y + \epsilon S(0; 1) \geqq 0.$$

We obtain

$$S(0; 1) = -S(0; 1) \leqq \frac{1}{\epsilon} y,$$

as was to be proved.

Lemma 7.2 gives [10]

LEMMA 7.3. *If E is E_B then every positive linear function $p(x)$ on E is N-continuous.*

THEOREM 7.1. *E_B is Archimedean if and only if E^+ is N-closed.*

PROOF. Suppose that E is $E_{A,B}$. Let y be an arbitrary point of the N-boundary of E^+ and u an N-inner point of E^+. Since E^+ is convex, every

$$\alpha y + (1 - \alpha)u$$

with $0 \leqq \alpha < 1$ belongs to E^+, i.e.

$$\alpha y + (1 - \alpha)u \geqq 0,$$

whence

$$y + \frac{1 - \alpha}{\alpha} u \geqq 0.$$

By Lemma 3.2 this inequality holds for $\alpha = 1$ which gives

$$y \geqq 0.$$

The sufficiency of the condition can be deduced from Lemma 3.2 in an elementary way.

THEOREM 7.2. *Let E be E_B. If E is weakly Archimedean then every point y of the N-boundary of E^+ belongs either to E^+ or to W.*

PROOF. It follows from Lemma 7.3 that W is N-closed in $E_{A^*,B}$ and consequently $\tilde{E} = E \ominus W$ is a Banach space if we put $\| \tilde{x} \| = \inf \{ \| x \|; x \in \tilde{x} \}$. It is also easy to see that $\tilde{E}$ is $\tilde{E}_B$.

If $\tilde{y} = y + W$ then $\tilde{y} \in \tilde{E}^+$ (use Theorems 4.2 and 7.1). If $\tilde{y} = \tilde{0} = W$ then $y \in W$; $\tilde{y} \neq \tilde{0}$ implies $y \bar{\in} W$, $y \in E^+$, as was to be proved.

If E is $E_{B,H}$ then the u-norm can be defined in E. According to Axiom B every N-bounded subset of E is bounded which implies that the N-topology in

E is generally stronger than the u-topology. They are obviously equivalent if, conversely, every bounded subset of E is N-bounded. Since every bounded subset $M \subset E$ can be included in an interval of the system $\mathfrak{A}$, we thus obtain

THEOREM 7.3. *If E is a homogeneous E_B then the N-norm in E is topologically equivalent to the u-norm if and only if the intervals in E are N-bounded.*

We say that the N-norm is *non-decreasing* if

$$-y \leqq x \leqq y$$

implies

$$\| x \| \leqq \| y \|.$$

LEMMA 7.4. *The N- and u-norms are topologically equivalent if and only if the N-norm is topologically equivalent to a non-decreasing norm (in particular, is non-decreasing itself).*

PROOF. Obvious, since the u-norm is non-decreasing and any two non-decreasing norms are topologically equivalent.

According to the results obtained by M. Krein ([10] and [11]) the N-boundedness of the intervals in E is equivalent to the following condition imposed on the adjoint space $\bar{E}$:

Every linear functional $f \in \bar{E}$ can be represented

$$f = h_1 - h_2 ,$$

where h_1, h_2 are positive linear functionals, i.e. $h_i(x) \geqq 0$, whenever $x \in E^+$ $(i = 1,2)$.

MATHEMATICAL INSTITUTE OF THE
ACADEMY OF SCIENCES OF THE U.S.S.R.

REFERENCES

[1] H. FREUDENTHAL, Proc. Kon. Akad. Amsterdam, XXXIX, No. 5 (1936).
[2] L. KANTOROVITCH, Rec. Math., 2(44): 1 (1937).
[3] B. L. VAN DER WAERDEN, Moderne Algebra, Berlin (1930).
[4] G. BIRKHOFF, Proc. Cambridge Phil. Soc., 29 (1933).
[5] A. TYCHONOFF, Math. Annalen, 111 (1935).
[6] A. YOUDINE, Annals of Leningrad University, 12 (1941).
[7] H. M. MACNEILLE, Proc. Nat. Acad. U.S.A., 22 (1936).
[8] M. KREIN, S. KREIN, C. R. de l'Acad. Sc. U.R.S.S., XXVII, N5 (1940).
[9] P. ALEXANDROFF, E. HOPF, Topologie, I, Berlin (1935).
[10] N. AKHIESER, M. KREIN, On some problems in the theory of moments (in Russian), Kharkhov, (1938).
[11] M. KREIN, C. R. de l'Acad. Sc. U.R.S.S., XXVIII, N1 (1940).

ANNALS OF MATHEMATICS
Vol. 44, No. 4, October, 1943

HOMOLOGY WITH LOCAL COEFFICIENTS

BY N. E. STEENROD

(Received January 26, 1942)

1. Introduction

In a recent paper [16] the author has had occasion to introduce and use what he believed to be a new type of homology theory, and he named it *homology with local coefficients*. It proved to be the natural and full generalization of the Whitney notion of *locally isomorphic complexes* [18]. Whitney, in turn, credits the source of his idea to de Rham's *homology groups of the second kind in a non-orientable manifold* [13]. It has since come to the author's attention that homology with local coefficients is equivalent in a complex to Reidemeister's *Überdeckung* [10].

Since this new homology theory (which includes the old) seems to have such wide applicability, a complete review of the older theory is needed to determine to what extent and in what form its theorems generalize. The object of this paper is to make such a survey. The general conclusion is that all major parts of the older theory do extend to the new. In addition the newer theory fills in several gaps in the old. The most noteworthy of these is a full duality and intersection theory in a non-orientable manifold (§14).

For the sake of completeness, some of the results of Reidemeister have been included. The new approach and new definitions make for easier and more intuitive proofs. They lead also to results not obtained by Reidemeister. The most important is a proof of the topological invariance of all the homology groups obtained.[1] In addition developments are given of the subjects of multiplications of cycles and cocycles, chain mappings, continuous cycles, and Čech cycles.

Part I contains an abstract development of systems of local groups in a space entirely apart from their applications to homology. Any fibre bundle over a base space R [18] determines many such systems in R (one for each homology group, homotopy group, etc., of the fibre). These are invariants of the bundle. They should prove to be of some help in classifying fibre bundles.

Part II, which contains the extended homology theory, presupposes on the part of the reader a knowledge of the classical theory such as can be found in the books of Lefschetz [7] and Alexandroff-Hopf [1].

I. LOCAL GROUPS IN A SPACE

2. Notations

We shall be dealing throughout with an arcwise connected topological space R. For any point x of R, let F_x be the fundamental (Poincaré) group of R with

[1] It is not determined however whether or not the combinatorial invariant called "torsion" by Reidemeister [9] and its generalizations by Franz [4] and de Rham [14] are true topological invariants.

x as initial and terminal point. If A is a curve from x to y, the class of curves from x to y homotopic to A with end points fixed we shall denote by a symbol such as α_{xy}. Its inverse is denoted by α_{xy}^{-1} or α_{yx}. The elements of F_x are abbreviated α_x, β_x, etc. The class α_{xy} determines an isomorphism $F_x \to F_y$ (denoted by α_{xy}) defined by $\alpha_{xy}(\beta_x) = \alpha_{yx}\beta_x\alpha_{xy}$. In keeping with this notation, the product $\alpha_x\beta_x$ means the element of F_x obtained by traversing first a curve of the class α_x then one of β_x. As is well known, the combination of two isomorphisms $\beta_{yz}(\alpha_{xy}(\gamma_x))$ is the isomorphism $(\alpha_{xy}\beta_{yz})(\gamma_x)$.

3. Local groups

We shall say that we have *a system of local groups (rings) in the space R* if (1) for each point x, there is given a group (ring) G_x, (2) for each class of paths α_{xy}, there is given a group (ring) isomorphism $G_x \to G_y$ (denoted by α_{xy}), and (3) the result of the isomorphism α_{xy} followed by β_{yz} is the isomorphism corresponding to the path $\alpha_{xy}\beta_{yz}$.

It follows from the transitivity condition (3) that the identity path from x to x is the identity transformation in G_x. A further consequence is that the inverse of the isomorphism α_{xy} is α_{yx}. By (2), a closed path $\alpha_x \in F_x$ determines an automorphism of G_x. From (3) it follows that F_x is a group of automorphisms of G_x. The invariant subgroup of F_x acting as the identity on G_x is denoted F_x^1. Since, by (3),

$$\alpha_{xy}(\beta_x(\alpha_{yx}(g))) = (\alpha_{yx}\beta_x\alpha_{xy})(g), \qquad g \in G_y,$$

it follows that

$$\alpha_{xy}(\beta_x(g)) = [\alpha_{xy}(\beta_x)](\alpha_{xy}(g)), \qquad g \in G_x. \tag{3.1}$$

We shall say that the system $\{G_x\}$ is *simple* if every $F_x^1 = F_x$. If this happens for one x, it will be true for all. If $\{G_x\}$ is simple, the isomorphism α_{xy} is independent of the path from x to y. Choosing a fixed point o as origin, we find that each G_x is uniquely isomorphic to G_o. Thus the local system consists of one G_o and as many copies of G_o as there are points $x \neq 0$.

Two systems $\{G_x\}$, $\{H_x\}$ are said to be *isomorphic* if, for each x, there is an isomorphism ϕ_x of G_x onto H_x such that

$$\phi_y(\alpha_{xy}(g)) = \alpha_{xy}(\phi_x(g)), \qquad g \in G_x.$$

We shall deal only with properties of systems which are invariant under isomorphisms. In each case the proof of invariance is trivial and will be omitted.

It was proved in §2 that the collection $\{F_x\}$ is a system of local groups. It is simple if and only if it is abelian.

In some instances a system $\{G_x\}$ will consist of topological groups. The isomorphisms α_{xy} will then be continuous. In the following pages we shall omit continuity considerations whenever such are reasonably obvious.

We shall consistently attempt to reduce the study of a system to the study of what occurs at one point of R. As a first step we have

THEOREM 1. *If G is a group (ring), o a point of R, and ψ a homomorphism of F_o into the group of automorphisms of G, then there is one and only one system $\{G_x\}$ of local groups (rings) in R such that G_o is a copy of G and the operations of F_o in G_o are those determined by ψ.*

For each point $x \in R$, choose a class of paths λ_{ox}, choosing λ_{oo} = identity. Let G_x be a group (ring) isomorphic to G. Associate this isomorphism with λ_{xo}. If α_{xy} is a path, we attach to it the isomorphism $G_x \to G_y$ defined by

$$\alpha_{xy}(g) = \lambda_{oy}[(\lambda_{ox}\alpha_{xy}\lambda_{yo})(\lambda_{xo}(g))].$$

The second operation is the automorphism of G attached to $\lambda_{ox}\alpha_{xy}\lambda_{yo} \in F_o$ by ψ. Since ψ is a homomorphism, the transitivity condition (3) holds.

If $\{G'_x\}$ is a local system, and ϕ an isomorphism of G'_o into G_o such that $\alpha\phi = \phi\alpha$ for $\alpha \in F_o$, map $G'_x \to G_x$ with the operation $\phi_x(g') = \lambda_{ox}(\phi(\lambda_{xo}(g')))$. It is easily verified that $\{\phi_x\}$ establishes an isomorphism between $\{G'_x\}$ and $\{G_x\}$. This proves the uniqueness and completes the theorem.

4. Automorphisms

Let $\{G_x\}$, $\{A_x\}$ be two systems of local groups, and suppose that A_x is a group of automorphisms of G_x in such a way that, for any α_{xy}, we have

$$\alpha_{xy}(a(g)) = \alpha_{xy}(a)(\alpha_{xy}(g)), \qquad a \in A_x,\ g \in G_x. \tag{4.1}$$

Then $\{A_x\}$ is called *a system of local automorphisms of* $\{G_x\}$. By (3.1), it follows that $\{F_x\}$ is such a system for any $\{G_x\}$.

Let A_x^1 be the subgroup of A_x acting as the identity on G_x. By (4.1), $\{A_x^1\}$ is a system of local groups. It follows that, under the natural isomorphisms $A_x/A_x^1 \to A_y/A_y^1$ induced by $A_x \to A_y$, $\{A_x/A_x^1\}$ is a system of local automorphisms of $\{G_x\}$.

If A is a group, and, for each x, A is a group of automorphisms of G_x such that

$$\alpha_{xy}(a(g)) = a(\alpha_{xy}(g)), \qquad a \in A,\ g \in G_x, \tag{4.2}$$

we shall call A *a group of uniform automorphisms of* $\{G_x\}$. If a system $\{A_x\}$ of local automorphisms is simple, then any one of its groups is, in a natural way, a group of uniform automorphisms of $\{G_x\}$. If F_x or F_x/F_x^1 is abelian, we have such a group of uniform automorphisms for any $\{G_x\}$.

As in Theorem 1, complete knowledge of a system of automorphisms is obtainable from knowledge of what occurs at a single point.

THEOREM 2. *Let A be a group of automorphisms of G with only the identity acting as such in G (i.e. $A^1 = 1$). Let o be a point of R, and let F_o be represented as a group of automorphisms of G in such a way that the automorphism $\alpha(a(\alpha^{-1}(g)))$ of G is in A for every $\alpha \in F_o$, $a \in A$. Then there is one and only one system $\{A_x\}$ of local automorphisms of $\{G_x\}$ such that the collection (F_o, G_o, A_o) is isomorphic to (F_o, G, A). $\{A_x\}$ is simple and therefore A is a group of uniform automorphisms of $\{G_x\}$ if and only if the automorphisms of F_o and A in G commute.*

By Theorem 1, the system $\{G_x\}$ is completely determined. By assumption

the automorphism $\alpha a \alpha^{-1}$ is a unique element of A. Thus each α determines an automorphism of A, and F_o is a group of such. By Theorem 1, the system $\{A_x\}$ of local groups is completely determined. For $a \in A_x$, $g \in G_x$, we choose a path α_{ox} and define

$$a(g) = \alpha_{ox}(\alpha_{xo}(a)(\alpha_{xo}(g))).$$

Clearly (4.1) will hold once we have proved the right side to be independent of α_{ox}. This is shown as follows.

$$\begin{aligned}\beta_{xo}(\alpha_{ox}(\alpha_{xo}(a)(\alpha_{xo}(g)))) &= (\alpha_{ox}\beta_{xo})(\alpha_{xo}(a)(\alpha_{xo}(g))) \\ &= [(\alpha_{ox}\beta_{xo})(\alpha_{xo}(a))(\alpha_{ox}\beta_{xo})^{-1}]((\alpha_{ox}\beta_{xo})(\alpha_{xo}(g))) \\ &= \beta_{xo}(a)(\beta_{xo}(g)).\end{aligned}$$

5. Operator rings

If G is an additive abelian group, the set H of all homomorphisms of G into itself forms a ring under the operations

$$(a + b)(g) = a(g) + b(g), \qquad (ab)(g) = a(b(g)), \qquad a, b \in H,\ g \in G.$$

A group A of automorphisms of G forms a multiplicative subgroup of H. It generates a subring A^* of H with unit = identity. If the symbol $a(g)$ be abbreviated by ag, this multiplication of $g \in G$ by the scalar $a \in A^*$ obeys the usual laws: $(a + b)g = ag + bg$, $(ab)g = a(bg)$, $1g = g$, $0g = 0$, $a(g + g') = ag + ag'$. One may therefore speak of linear combinations, linear independence, and bases in G relative to A.

We shall say that the system $\{A_x\}$ is *a system of operator rings for the abelian system* $\{G_x\}$ if, for each x, A_x is a ring of operators for G_x, and for each path α_{xy}, we have

$$\alpha_{xy}(ag) = \alpha_{xy}(a)\alpha_{xy}(g), \qquad a \in A_x,\ g \in G_x.$$

The analogue of Theorem 2 is proved with only slight modifications. If a system of operator rings is simple, we are led naturally to the concept of *a uniform operator ring for* $\{G_x\}$.

For any $\{G_x\}$, the system $\{F_x^*\}$ of group rings of $\{F_x\}$ is a system of operator rings. If F_x or F_x/F_x^1 is abelian, the group ring of F_o or factor ring thereof is a uniform operator ring for $\{G_x\}$.

6. Dual systems

Two abelian systems $\{G_x\}$, $\{H_x\}$ of local groups form a *pair* with respect to a third $\{K_x\}$ if, for each x, G_x and H_x form a pair with respect to K_x (i.e. a multiplication $gh = k$ is given which is linear in each factor) in such a way that, for each path α_{xy}, we have $\alpha_{xy}(gh) = \alpha_{xy}(g)\alpha_{xy}(h)$. Analogous to Theorem 1, we have

THEOREM 3. *Let G, H form a pair with respect to K, and let F_o be realized as a group of automorphisms of each of G, H, K in such a way that $\alpha(gh) = \alpha(g)\alpha(h)$ for $\alpha \in F_o$, $g \in G$, $h \in H$. Then there is one and only one set of systems $\{G_x\}$, $\{H_x\}$,*

$\{K_x\}$ such that the first two form a pair with respect to the third, and G_o, H_o, K_o and the automorphisms F_o form a set isomorphic to the given G, H, K, F_o.

By Theorem 1, there are unique systems $\{G_x\}$, $\{H_x\}$, $\{K_x\}$ corresponding to G, H, K and the operations of F_o in these groups. In the construction of all three systems let us use the same collection of paths λ_{ox} (see proof of Theorem 1). Under the isomorphism of G_x, H_x, K_x with G, H, K attached to λ_{ox} the multiplication in the latter groups carries over into a multiplication in the former. The relation $\alpha_{xy}(gh) = \alpha_{xy}(g)\alpha_{xy}(h)$ for $g \in G_x$, $h \in H_x$ is proved by using the property $\alpha(g'h') = \alpha(g')\alpha(h')$ of the closed path $\alpha = \lambda_{ox}\alpha_{xy}\lambda_{yo}$ where g', h' correspond to g, h under λ_{ox}. Any other allowable product which agrees with the constructed product in G_o, H_o will likewise agree in G_x, H_x since both products are invariant under the translation along λ_{xo}.

Of particular interest is the case of character groups.

THEOREM 4. *If K = real numbers* mod 1, *G = a discrete, or compact, or locally compact separable group, and F_o is realized in two arbitrary ways as a group of automorphisms of K_o and G_o respectively, then F_o can be realized in one and only one way as a group of automorphisms of the character group H of G satisfying $\alpha(gh) = \alpha(g)\alpha(h)$. Thus to given systems $\{K_x\}$, $\{G_x\}$ is attached a unique system $\{H_x\}$ of character groups.*

The character $\alpha(h)$ is defined to be the one with the value $\alpha(\alpha^{-1}(g)h)$ on any g. The remainder of the theorem is readily verified.

It is to be noted that K admits but one non-trivial automorphism, namely: $k \to -k$. Thus there are as many character systems of a given system $\{G_x\}$ as there are factor groups of F_o of order 2. If it should be desirable to have a *unique* character system, it would be natural to choose $\{K_x\}$ to be simple.

7. Local groups under mappings

Let R, R' be two arcwise connected spaces and ϕ a continuous map $R' \to R$. Let F, F' be their fundamental groups relative to base points o, o' such that $\phi(o') = o$. If $\{G_x\}$ is a system of local groups in R, then ϕ *induces* in R' a system $\{G'_y\}$ as follows. The group G'_y is chosen isomorphic to G_x where $x = \phi(y)$. The isomorphism is denoted by ϕ. The path α_{yz} in R' maps G'_y isomorphically on G'_z according to the rule

$$\alpha_{yz}(g) = \phi^{-1}(\phi(\alpha_{yz})(\phi(g))), \qquad g \in G'_y. \tag{7.1}$$

The transitivity condition is immediately verified.

The existence of the induced system is apparent in view of Theorem 1 and the fact that the homomorphism $F' \to F$ realizes F' as a group of automorphisms of G_o.

The induced system has numerous properties which we list without proofs.

(a) If $\{G_x\}$ is simple, so is $\{G'_y\}$.

(b). $\{G'_y\}$ is simple if and only if F' is mapped by ϕ on the subgroup F^1 of F leaving G_o fixed.

(c). If $\{G_x\}$, $\{H_x\}$ are paired relative to $\{K_x\}$, then likewise the induced systems in R'.

(d). If $\{A_x\}$ is a system of local automorphisms or operator rings for $\{G_x\}$, then likewise the induced systems in R'.

(e). If A is a group of uniform automorphisms or a uniform operator ring for $\{G_x\}$, it is also one for the induced system.

It follows from (b) that, if R' is the covering space of R corresponding to the subgroup F^1 of F, the induced system is simple. Thus any system in R can be considered as the continuous image of a simple system in some covering space.

It is natural to inquire under what circumstances a given system in R' is induced by one in R. For this it is necessary and sufficient that (1) the kernel $\bar{F}'$ of the homomorphism $F' \rightarrow F$ shall act as the identity on $G'_{o'}$, and (2) the map of the subgroup $F'/\bar{F}'$ of F into the group of automorphisms of $G'_{o'}$ shall admit a homomorphic extension to F.

8. Special local groups

Because of their importance in the work of Reidemeister, we shall discuss certain local systems based on the fundamental group F of R.

Let F^1 be an invariant subgroup of F, and let $\mathfrak{F} = F/F^1$. Let $\mathfrak{G}$ be an abelian group. Let G be the set of functions from $\mathfrak{F}$ to $\mathfrak{G}$. If two such functions are added by adding functional values at each element of $\mathfrak{F}$, G becomes an abelian group. A function $f \in G$ is called a *restricted* function if $f(\alpha) = 0$ except for a finite number of $\alpha \in \mathfrak{F}$. The restricted functions form a subgroup G' of G. The structure of G (G') can be described as the unrestricted (restricted) direct sum of as many copies of $\mathfrak{G}$ as there are elements of $\mathfrak{F}$. If $\mathfrak{G}$ = integers and $F^1 = 1$, then G' is the ordinary group ring of F.

If $\mathfrak{G}$ is a ring, we define a multiplication of two functions f, $g \in G'$ by

$$f \times g(\alpha) = \textstyle\sum_{\beta} f(\alpha\beta^{-1})g(\beta), \qquad \alpha, \beta \in \mathfrak{F}.$$

Since the functions in G' are restricted, the sum is finite; thus G' is a ring. If $\mathfrak{G}$ = integers, this product is the usual one in the group ring.

The group F can be realized in three natural ways as a group of automorphisms of the groups G, G'. For any $\gamma \in F$, we define three operations on a function $f \in G$ (G') by

$$L_\gamma f(\alpha) = f(\gamma^{-1}\alpha), \qquad R_\gamma f(\alpha) = f(\alpha\gamma), \qquad T_\gamma f(\alpha) = f(\gamma^{-1}\alpha\gamma), \qquad \alpha \in \mathfrak{F}.$$

Then $L_\gamma f$, $R_\gamma f$ and $T_\gamma f$ are in G (G'), and are called the *left translation*, *right translation*, and *transform* of f by γ, respectively. It is easy to verify that L_γ, R_γ, T_γ are automorphisms of G (G'), and that $L_\gamma L_\delta = L_{\gamma\delta}$, $R_\gamma R_\delta = R_{\gamma\delta}$, and $T_\gamma T_\delta = T_{\gamma\delta}$. The subgroup acting as the identity for both L and R is F^1.

In the case that $\mathfrak{G}$ and therefore G' is a ring, the left and right translations are not *ring* automorphisms. However the transforms are: $T_\gamma(f \times g) = (T_\gamma f) \times (T_\gamma g)$.

It follows from Theorem 1 that, corresponding to each of the three ways that

F can act on G (G'), we have a system of local groups. These we denote by $\{G_x^L\}$, $\{G_x^R\}$, $\{G_x^T\}$, and similarly for G'. The last of these, $\{G_x'^T\}$, is a system of local rings, whenever $\mathfrak{G}$ is a ring.

Since $\mathfrak{G}$ is associative, $L_\gamma R_\beta f = R_\beta L_\gamma f$. Therefore, by Theorem 2, the left translations form a group of uniform automorphisms of $\{G_x^R\}$, and likewise for the right translations of $\{G_x^L\}$. Then, as in §5, the group ring of F is a ring of uniform operators for these systems.

Under the automorphism ϕ of G (G') defined by $\phi f(\alpha) = f(\alpha^{-1})$, we have $\phi L_\gamma f = R_\gamma \phi f$. Therefore, by means of ϕ, an isomorphism exists between $\{G_x^L\}$ and $\{G_x^R\}$.

In the work of Reidemeister on homotopy chains, $\mathfrak{G}$ = integers, $F^1 = 1$, so that G' is the group ring of F. The coefficients of the homotopy chains belong to the local groups $\{G_x'^R\}$ (these are not rings), and the ring of uniform operators is likewise G' where left translations are used. This distinction between the two usages of the group ring of F is necessary for a comprehension of the subject of homotopy chains.

II. The combinatorial theory

9. Chains with local coefficients

Let $\{G_x\}$ be a system of local abelian groups in a space R which is decomposed into a cell complex[2] K. A q-cell of K is denoted by σ^q, incidence by $\sigma < \sigma'$, and incidence numbers by $[\sigma^{q-1}:\sigma^q]$. We suppose as usual that

$$\sum_{\sigma^{q-1}} [\sigma^{q-2}:\sigma^{q-1}][\sigma^{q-1}:\sigma^q] = 0. \tag{9.1}$$

In each cell σ we choose a representative point $x(\sigma)$ and abbreviate the symbol $G_{x(\sigma)}$ by G_σ. A q-chain of K is a function[3] f attaching to each oriented q-cell σ an element $f(\sigma) \in G_\sigma$ with the property $f(-\sigma) = -f(\sigma)$. Chains are added by adding functional values. They then form a group isomorphic to the direct sum of the groups G_σ for all q-cells σ.

If $\sigma' < \sigma$, we may choose a path in the closure of σ joining $x(\sigma)$ to $x(\sigma')$ and obtain therefrom an isomorphism $G_\sigma \to G_{\sigma'}$ which is denoted by $h_{\sigma'\sigma}$. In order that $h_{\sigma'\sigma}$ shall be independent of the path, we postulate that the closure of each cell is simply connected. A second consequence of this and the transitivity condition is

$$h_{\sigma''\sigma'} h_{\sigma'\sigma} = h_{\sigma''\sigma} \tag{9.2}$$

[2] For the sake of simplicity we suppose K is finite and closed. The extension of the subsequent results to relative complexes, open complexes, and to the finite and infinite chains of locally-finite complexes will be obvious.

[3] We shall abide by the functional notation throughout. This will prove to be as convenient as the classical linear form notation. We abandon the latter since it has algebraic implications more prejudicial than suggestive in the present discussion.

By means of the isomorphisms h, we can define the boundary ∂f and coboundary δf of a q-chain f:

$$\begin{aligned} \partial f(\sigma^{q-1}) &= \textstyle\sum_{\sigma^q} [\sigma^{q-1}:\sigma^q] h_{\sigma^{q-1}\sigma^q}(f(\sigma^q)), \\ \delta f(\sigma^{q+1}) &= \textstyle\sum_{\sigma^q} [\sigma^q:\sigma^{q+1}] h^{-1}_{\sigma^q\sigma^{q+1}}(f(\sigma^q)). \end{aligned} \tag{9.3}$$

The sums extend over the q-cells which (a) have σ^{q-1} as a face, (b) are faces of σ^{q+1}. These new functions are $q - 1$ and $(q + 1)$-chains respectively. It follows from (9.1) and (9.2) that $\partial\partial f = 0$, $\delta\delta f = 0$. Therefore cycles, bounding cycles, homology and cohomology groups can be defined as usual.

A special convention for 0-cycles is necessary. We shall agree that any 0-chain is a 0-cycle. Note that the Kronecker index (i.e. the sum of the coefficients of a 0-chain) has no meaning unless the system $\{G_x\}$ is simple.[4]

Finally it is to be remarked that we obtain a completely isomorphic situation if we choose new representatives $y(\sigma)$ in each σ. The isomorphism is established by means of a path $x(\sigma)$ to $y(\sigma)$ in σ for each σ.

10. Automorphisms of chains

Let A be a group of *uniform* automorphisms or a ring of *uniform* operators of $\{G_x\}$. For any chain f and $a \in A$, define $(af)(\sigma) = af(\sigma)$ for each σ. Then af is a chain, and A appears as a group of automorphisms or a ring of operators of the group of q-chains for each q.

Since the operations of A commute with translations of the G_x along paths, it follows that the operations of A on the chains commute with ∂ and δ. Therefore A appears as an automorphism group or ring of operators of the groups of cycles, cocycles, boundaries, coboundaries, and consequently of the homology and cohomology groups.

11. Multiplication of chains

It should be noted that a cycle with local coefficients is locally a cycle in the ordinary sense; for, in any simply connected open set U (e.g. the star of a vertex), isomorphisms of the local groups can be set up with a fixed group (using paths in U) in such a way as to transform each chain with local coefficients into an ordinary chain mod $(K - U)$ so that the boundary relations are preserved. It follows that any operation on ordinary chains which is a sum of local operations can be carried over to chains with local coefficients. We have seen that this is true of the boundary operator. We shall see that this is also true of products of cocycles, products of cycles and cocycles, and intersections of cycles. The linking number like the Kronecker index is not of this category.

[4] Classical homology with a single coefficient group G is isomorphic to homology with coefficients in the simple system of local groups determined by G.

In a comprehensive paper of Whitney on products [17], it is proved that, corresponding to cells σ_i^p, σ_j^q, σ_k^{p+q} in K, there is an integer ${}^{pq}\Gamma_k^{ij}$ such that

(Γ_1) if σ_i^p, σ_j^q are not both faces of σ_k^{p+q}, then ${}^{pq}\Gamma_k^{ij} = 0$,
(Γ_2) for all p, q, i, j, k,

$$\sum_m [\sigma_m^{p+q}:\sigma_k^{p+q+1}]^{pq}\Gamma_k^{ij} = \sum_m [\sigma_i^p:\sigma_m^{p+1}]^{p+1,q}\Gamma_k^{mj} + (-1)^p \sum_m [\sigma_j^q:\sigma_m^{q+1}]^{p,q+1}\Gamma_k^{im},$$

(Γ_3) for all q and j, $\sum_i {}^{0q}\Gamma_j^{ij} = 1$.
Although the last two conditions appear not to be local in nature, they are so by virtue of the first.

Let $\{G_x\}$ be a system of local rings with units. If f^p and f^q are p and q-chains respectively with coefficients in $\{G_x\}$, we define their *cup* product to be the $(p + q)$-chain

$$(11.1) \qquad f^p \cup f^q(\sigma_k^{p+q}) = \sum_{ij} {}^{pq}\Gamma_k^{ij}[h_{k\ ,i}^{p+q,p}(f^p(\sigma_i^p))][h_{k\ ,j}^{p+q,q}(f^q(\sigma_j^q))].$$

Here we heve abbreviated the isomorphism of the local group of σ_i^p on that of σ_k^{p+q} by $h_{k\ ,i}^{p+q,p}$. The sum extends over the faces σ_i^p, σ_j^q of σ_k^{p+q}. The relations

(P_1) $f^p \cup f^q$ is zero on any σ^{p+q} which has not both a face in f^p and a face in f^q,
(P_2) $\delta(f^p \cup f^q) = \delta f^p \cup f^q + (-1)^p f^p \cup \delta f^q$,
(P_3) $I \cup f^q = f^q$,

follow from the relations (Γ), the transitivity of the h's (9.2), and the preservation of the products in G_x under an h. In (P_3), I is the 0-cocycle which attaches to each vertex V the unit of G_V. Since h is a ring isomorphism, it preserves the unit; therefore $\delta I = 0$. It follows just as in Whitney [17] that a product is definable for the cohomology classes with the usual properties. The associative law and the commutation rule for the special products in a simplicial complex are given local proofs. Once invariance under subdivision has been established (see §16), the same laws will hold in a general complex.

For the *cap* product, we shall suppose that $\{G_x\}$, $\{H_x\}$ are paired with respect to $\{L_x\}$. Given a q-cochain f^q (coef. $\{G_x\}$) and a $(p + q)$-chain g^{p+q} (coef. $\{H_x\}$), we define their cap product to be the p-chain

$$(11.2) \qquad f^q \cap g^{p+q}(\sigma_m^p) = \sum_{jk} {}^{pq}\Gamma_k^{mj}[h_{m,j}^{p,q}(f^q(\sigma_j^q))][h_{m,k}^{p,p+q}(g^{p+q}(\sigma_k^{p+q}))]$$

with coefficients in $\{L_x\}$. The sum extends over those j, k for which σ_m^p, σ_j^q are faces of σ_k^{p+q}. Just as before, we obtain

(Q_1) $f^q \cap g^{p+q}$ is zero on any σ^p which with no σ^q of f^q is a face of a σ^{p+q} of g^{p+q},

(Q_2) $\partial(f^q \cap g^{p+q}) = (-1)^p \delta f^q \cap g^{p+q} + f^q \cap \partial g^{p+q}$.

In order to obtain an analogue of (Γ_3), we shall take the special case where $\{L_x\} = \{H_x\}$ and $\{G_x\}$ is a system of operator rings for $\{H_x\}$ (see §5). Then the 0-cocycle I is defined, and we have (Q_3) $I \cap g^p = g^p$.

Under the same assumption, it can be shown that

$$(f^q \cup f^r) \cap g^{p+q+r} \sim f^q \cap (f^r \cap g^{p+q+r}),$$

for cocycles f^q, f^r and a cycle g^{p+q+r}.

The uniqueness of the products in the following sense can be proved. Suppose $^{pq}\bar{\Gamma}_k^{ij}$ is another set of Γ's for which the relations (Γ) hold. Corresponding to the two sets of Γ's is the Whitney operation $\wedge$ (Theorem 8, [17]). Let $^{pq}\Delta_k^{ij}$ be the coefficient of σ_k^{p+1} in the product $\sigma_i^q \wedge \sigma_j^{p+q}$. Define $f^q \wedge g^{p+q}$ with equations analogous to (11.2) using Δ in place of Γ. Then the Whitney relations (R) (loc. cit.) with chains in place of cells can be proved to hold. It follows that the two sets of Γ's determine the same products among the homology and cohomology classes.

12. Duality

Let L be the group of real numbers mod 1, and let $\{L_x\}$ be the corresponding simple system in K. Let $\{G_x\}$ and $\{H_x\}$ be character systems of one another with respect to $\{L_x\}$. Since $\{L_x\}$ is simple, a 0-cycle f^0 with coefficients in $\{L_x\}$ may be regarded as a 0-cycle with coefficients in L. It possesses therefore a Kronecker index (= the sum of its coefficients) which we denote by (f^0). As usual $(f^0) = 0$ is equivalent to $f^0 \sim 0$. The scalar product of a q-cochain f^q coef. $\{G_x\}$ and a q-chain g^q coef. $\{H_x\}$ is defined to be the Kronecker index of their cap product:

(12.1) $$f^q \cdot g^q = (f^q \cap g^q) \quad \text{in } L.$$

If f^q and g^q are zero except on a single σ, then, by Γ_3, $f^q \cdot g^q$ is the product $f^q(\sigma)g^q(\sigma)$. Therefore, due to the linearity, the scalar product of arbitrary f^q, g^q is the sum of products of corresponding coefficients. It follows that the groups of q-cochains and q-chains are character groups of one another.

For any q-cochain f^q and $(q + 1)$-chain g^{q+1} we obtain from Q_2 that

(12.2) $$\delta f^q \cdot g^{p+1} = f^q \cdot \partial g^{q+1}.$$

It follows now in the usual way (see Whitney [17]) that the q^{th} cohomology group coef. $\{G_x\}$ and the q^{th} homology group coef. $\{H_x\}$ are character groups with the scalar product as the multiplication.

13. Intersection in an orientable manifold

Let K be an orientable simplicial n-manifold and let K^* be its dual. Denote by $\mathfrak{D}\sigma$ the cell of K^* dual to the oriented simplex σ of K relative to a fixed choice of the fundamental n-cycle Z^n with integer coefficients. Let $\{G_x\}$ be a system of local groups to be used as coefficients in both K and K^*. Let the coefficient groups of σ and $\mathfrak{D}\sigma$ be the group G_x where x is their common point. Then for any chain f of K (cochain f^* of K^*), the equation

(13.1) $$f^*(\mathfrak{D}\sigma) = f(\sigma)$$

defines its *dual* cochain (chain) of K^* (K) of the complementary dimension. This isomorphism between the two groups of chains has, as usual, the property

$$(13.2) \qquad (\partial f)^* = (-1)^q \delta f^*, \qquad q = \text{dimension of } f.$$

It follows that *the* q^{th} *homology group coef.* $\{G_x\}$ *and the* $(n - q)^{\text{th}}$ *cohomology group coef.* $\{G_x\}$ *are isomorphic.*

In case $\{G_x\}$ is a system of local rings, we have as in §11 a multiplication defined for the cochains of K^*. The isomorphisms just established between the chains of K and cochains of K^* enable us to carry over the product in K^* into an intersection in K. We define the *intersection* of two chains f_1, f_2 of K to be the dual of the cup product of their duals:

$$(13.3) \qquad f_1 \circ f_2 = (f_1^* \cup f_2^*)^*.$$

(Compare Whitney [17], p. 422, formula (19.9)). It follows that *a system of local rings in an orientable manifold determines an intersection ring of cycles isomorphic to the ring of cocycles (same coefficients) under the operation of dual.*

As is well known, the ring of integers is a ring of operators for any group G which commute with any automorphism of G. The ring of integers is therefore a uniform operator ring (§5) for any $\{G_x\}$. The equation (11.1) may be interpreted as defining the cup product of a cochain f^p with (simple) integer coefficients and a cochain f^q with local coefficients $\{G_x\}$. The relations (P) still hold, and these together with the associative law lead to the conclusion that *the cohomology ring with simple integer coefficients constitutes a ring of operators for the cohomology groups coef.* $\{G_x\}$.

The dual of this last result is that *the intersection ring of an orientable manifold with simple integer coefficients is a ring of operators for the homology groups coef.* $\{G_x\}$.

14. Intersection in a non-orientable manifold

In a non-orientable manifold K there is no n-cycle with simple integer coefficients. One cannot therefore determine the orientation of $\mathfrak{D}\sigma$ uniformly over K so that (13.2) holds. A customary device is to use integers mod 2 as coefficients so as to restore the basic n-cycle and escape orientation difficulties. The resulting duality and intersection theory is a bit weak due to the inadequacy of the coefficients. A more ingenious device has been used by de Rham [13]. We shall see that a suitable use of local coefficients permits a full development of De Rham's notion and leads to a complete and satisfying duality and intersection theory in a non-orientable manifold.

Since K is non-orientable, the elements of its fundamental group F divide into two classes according as they do or do not preserve orientation. Those which do form an invariant subgroup F^1 of index 2. Let T be the group of integers. For each integer $t \in T$ and $\alpha \in F$, let $\alpha(t)$ be $+t$ or $-t$ according as α is or is not in F^1. In this way F is a group of automorphisms of T. Let $\{T'_x\}$ be the corresponding system of local groups given by Theorem 1. We shall say that chains with coefficients in $\{T'_x\}$ have *twisted integer coefficients.*

Let G be an abelian group and let F be represented as a group of automorphisms of G. Let $\bar{G}$ be the direct sum of two copies of G (i.e. the group of pairs (g_1, g_2)). Identify G with the subgroup of elements of the form $(g, 0)$, and call G the real part of $\bar{G}$. The subgroup G' of elements of the form $(0, g)$ we call the imaginary part of $\bar{G}$. The product of (g_1, g_2) with the complex number $a + ib$ (a, b are integers) is defined by

$$(14.1) \qquad (a + ib)(g_1, g_2) = (ag_1 - bg_2, ag_2 + bg_1).$$

It follows immediately that the complex integers form a ring of operators for the group $\bar{G}$, and that each element of $\bar{G}$ can be written uniquely in the form $g_1 + ig_2$ (g_1, g_2 real). If we set

$$(14.2) \qquad \alpha(g_1 + ig_2) = \begin{cases} \alpha(g_1) + i\alpha(g_2) & \text{if } \alpha \in F^1, \\ \alpha(g_1) - i\alpha(g_2) & \text{if } \alpha \notin F^1, \end{cases}$$

the automorphisms of F in G are extended to $\bar{G}$. For any complex integer $a + ib$, define $\alpha(a + ib) = a \pm ib$ according as α is or is not in F^1. Let $\bar{T}$ be the ring of complex integers, and let $\{\bar{T}_x\}$ be the system of local rings corresponding to $\bar{T}$ and these automorphisms. It follows that $\{\bar{T}_x\}$ is a system of operator rings for the system $\{\bar{G}_x\}$ corresponding to $\bar{G}$ and the automorphisms (14.2) (see §5). We refer to $\{\bar{G}_x\}$ as the *complex extension of* $\{G_x\}$.

If G is a ring, and the elements of F are ring automorphisms, we define a product in $\bar{G}$ in the usual way: $(g_1, g_2)(g_1', g_2') = (g_1g_1' - g_2g_2', g_2g_1' + g_1g_2')$. The equations (14.2) define ring automorphisms of $\bar{G}$. Furthermore the elements of $\bar{T}$ associate and commute with the multiplications in $\bar{G}$. Thus $\{\bar{G}_x\}$ is a system of local rings with $\{\bar{T}_x\}$ as a system of operator rings.

We shall use $\{\bar{G}_x\}$ and $\{\bar{T}_x\}$ as coefficients for chains of K and cochains of K^*. The group of n-cycles of K coef. $\{\bar{T}_x\}$ is infinite cyclic. A generator is constructed as follows. Choose an oriented n-cell σ and let $Z^n(\sigma) = \pm i$ in $\bar{T}_\sigma$ (because of $\alpha(i) = \pm i$, the sign of i has only a local significance). If $|\sigma'|$ is any other n-cell, choose a path α of n-cells from σ to $|\sigma'|$ (successive cells having an $(n - 1)$-face in common). The path α determines an orientation σ' of $|\sigma'|$ concordant with that of σ. Now define $Z^n(\sigma') = \alpha(Z^n(\sigma))$. In words: the orientation and coefficient of σ' are determined by translating along a path α both the orientation and coefficient of σ. Translating along a second path β will either produce the same orientation and the same coefficient or reverse the sign of both according as $\alpha\beta^{-1}$ is or is not in F^1. Thus the chain Z^n is independent of the paths chosen in its construction. It is a cycle since, in any simply connected domain, it is a cycle. The usual argument shows that any n-cycle of K coef. $\{\bar{T}_x\}$ is an integral multiple of Z^n. Thus the fundamental n-cycle of K has pure imaginary coefficients (or equally well, twisted integer coefficients).

If ϵ is an oriented simply-connected neighborhood in K, $Z^n(\epsilon)$ will be the coefficient $Z^n(\sigma)$ of an n-cell σ in ϵ oriented concordantly with ϵ. Corresponding to a q-cell $\sigma \in K$ and an oriented neighborhood ϵ of σ, there is in the usual way a

unique orientation of the dual cell $\mathfrak{D}_\epsilon\sigma$ in K^*. If f is any q-chain of K coef. $\{\bar{G}_x\}$, we define its dual to be the $(n - q)$-cochain f^* in K^* coef. $\{\bar{G}_x\}$ defined by

$$(14.3) \qquad f^*(\mathfrak{D}_\epsilon\sigma) = -f(\sigma)Z^n(\epsilon).$$

It is to be understood that $f(\sigma)$ is in $\bar{G}_x$ and $Z^n(\epsilon)$ is in $\bar{T}_x$ where x is the point common to σ and $\mathfrak{D}_\epsilon\sigma$. Since reversing the orientation of ϵ changes the sign of both sides of (14.3), f^* is independent of the choices of the ϵ's. The dual of a chain f^* in K^* is given by

$$(14.4) \qquad f^{**}(\sigma) = f^*(\mathfrak{D}_\epsilon\sigma)Z^n(\epsilon).$$

Since $Z^n(\epsilon)Z^n(\epsilon) = -1$, we have that *any chain is the dual of its dual.* The dual of Z^n is the unit 0-cocycle I. The dual of a real (imaginary) chain is imaginary (real). The formula (13.2) now follows for the dual in a non-orientable manifold; for it is a statement of local properties, and (14.3), (14.4) differ from (13.1) locally by a constant factor. As in §13, it follows that *the* q^{th} *homology group* coef. $\{\bar{G}_x\}$ *and the* $(n - q)^{\text{th}}$ *cohomology group* coef. $\{\bar{G}_x\}$ *are isomorphic.* Using (13.3) to define the intersection, we obtain a homology ring isomorphic to the cohomology ring whenever the coef. $\{\bar{G}_x\}$ form a system of local rings. In any case, the homology ring coef. $\{\bar{T}_x\}$ forms a ring of operators for the homology groups coef. $\{\bar{G}_x\}$.

Since $\bar{G}$ is the direct sum of its real and imaginary part, any chain is uniquely a sum of a real chain and an imaginary chain. The operations ∂ and δ preserve the property of being real or imaginary. Therefore the homology and cohomology groups decompose into direct sums of their real and imaginary parts. Since passing to the dual interchanges real and imaginary, we obtain the following results. *The* q^{th} *homology group* coef. $\{G_x\}$ (coef. $\{G'_x\}$) *is isomorphic to the* $(n - q)^{\text{th}}$ *cohomology group* coef. $\{G'_x\}$ (coef. $\{G_x\}$). *The homology classes* coef. $\{G'_x\}$ *(i.e. the imaginary ones) form a ring isomorphic to the cohomology ring* coef. $\{G_x\}$. *The intersection of two real cycles is imaginary. The intersection of a real and an imaginary cycle is real.*

It is to be noted that the results of this section apply to an orientable manifold. The absence of orientation reversing paths in no wise invalidates the constructions. We have in this way a single theory including both types of manifolds.

The classical approach to intersection is to define directly the intersection of a p-chain f of K with a q-chain g of K^* to be a chain of the subdivision of K. We may do this here as follows. If σ is a p-simplex, σ' an $(n - q)$-face of σ and ϵ an oriented neighborhood of σ, define $\sigma \circ \mathfrak{D}_\epsilon\sigma'$ relative to ϵ in the usual way. Then define the intersection chain $f \circ g$ to have on $\sigma \circ \mathfrak{D}_\epsilon\sigma'$ the value $-f(\sigma)g(\mathfrak{D}_\epsilon\sigma')Z^n(\epsilon)$.

15. The Poincaré duality

If we assume that $\{G_x\}$, $\{H_x\}$ are character systems of one another with respect to the simple system $\{L_x\}$ of mod 1 groups, we may combine the results of §12 with those of §13 and §14 to obtain: *The* q^{th} *homology group of the manifold* K

coef. $\{G_x\}$ (coef. $\{G'_x\}$) *and the* $(n - q)^{\text{th}}$ *homology group* coef. $\{H'_x\}$ (coef. $\{H_x\}$) *are character groups of one another; the multiplication is determined by the scalar product* $(f^* \cap g)$ *where* f *is a* q*-chain of* K *and* g *is an* $(n - q)$*-chain of* K^*.

The results simplify in the orientable case if we note that $\{G_x\}$ and $\{G'_x\}$ are isomorphic, as also are $\{H_x\}$ and $\{G'_x\}$. In the non-orientable case, the results simplify if we note that $\{H'_x\}$ ($\{H_x\}$) is the character system of $\{G_x\}$ ($\{G'_x\}$) with respect to the system $\{L'_x\}$ of twisted mod 1 groups associated with the simple system $\{L_x\}$.

Before leaving the subject of duality, let us observe that the notion of local coefficients has nothing to add to the duality theorem of Alexander. A vital step in the argument of Alexander is the following: *A cycle in the closed set* R *in the* n*-sphere* S^n *is a cycle in* S^n *and is therefore the boundary of a chain in* S^n. Such a statement is valid only if the system of local coefficients used in R is part of a system in S^n; as S^n is simply-connected $(n > 1)$, the system of local coefficients must be simple.

16. Chain mapping and subdivision

A chain transformation is a homomorphism of the groups of chains of one complex on those of another which commutes with the boundary or coboundary operator or interchanges them (as in the case of the dual). This definition has meaning of course when local coefficients are used. All that we need to determine here is that chain transformation "S" with local coefficients exist in the usual circumstances.

Let a cell mapping $K' \to K$ be given which preserves the relation of incidence. Let $\{G_x\}$ be a system of local coefficients in K, and let $\{G'_y\}$ be the induced system in K' (see §7). If $\sigma' \to \sigma$, there is an attached isomorphism $G'_{\sigma'} \to G_\sigma$. A chain of K' which is zero except on a single σ' we call an *elementary chain*. Let f' be an elementary chain and $f'(\sigma') \neq 0$. If the image σ of σ' has a lower dimension, we define the image f of f' to be zero. If σ, σ' have the same dimension, the image f of f' is zero except on σ, and $f(\sigma)$ is the image of $f'(\sigma')$ under the isomorphism $G'_{\sigma'} \to G_\sigma$. An arbitrary chain of K' is uniquely a sum of elementary chains. Its image is defined as the sum of the images of its elementary parts. The resulting chain mapping we denote by τ. The inverse cochain mapping τ' attaches to an elementary chain of K the sum of the elementary chains of K' mapped into it by τ; we then extend τ' preserving linearity. That τ (τ') commutes with ∂ (δ) is proved by first establishing it in the usual way for elementary chains (of course (7.1) is used), and then applying the linearity of ∂ (δ).

It is necessary to use in K' the induced system $\{G'_y\}$ in order that τ, τ' shall exist in all dimensions. This is seen as follows. Let V' be a vertex of K' and V its image. Assuming τ, τ' defined for the elementary chains of V, V^1, we arrive at an isomorphism $G'_{V'} \to G_V$. Therefore $\{G'_y\}$ is the system induced by $\{G_x\}$ over the 0-dimensional part of K'. Suppose this is known for the q-dimensional part of K'. Any closed $(q + 1)$-cell is simply-connected, there is therefore just one system of local groups defined over it which agrees with a given

system on its boundary, and that one is simple. We conclude that the given system and the induced system agree on each closed $(q + 1)$-cell, and finally over the whole of K'.

If K' is a subdivision of the simplicial complex K, we then have two systems in K': the given system $\{G_x\}$ for K, and the system $\{G'_x\}$ induced by the map $K' \rightarrow K$ defined by mapping each vertex of K' into a vertex of the simplex of K containing it. The two systems are isomorphic. The isomorphism is set up by using the line segments which join each point to its image point. The proof of the invariance under subdivision of the groups of K and their multiplications may now be completed in the standard way (see for example [17]).

17. Continuous cycles

Let $\{G_x\}$ be a system of local groups in a space R. A *continuous chain* in R is a collection composed of a complex K, a continuous map ϕ of K in R, and a chain Z in K with local coefficients in the system $\{G'_y\}$ induced by ϕ and $\{G_x\}$. If Z is a cycle, the collection (K, ϕ, Z) is called a *continuous cycle.* The boundary of (K, ϕ, Z) is $(K, \phi, \partial Z)$. Two continuous chains (K_i, ϕ_i, Z_i) $(i = 1, 2)$ are *added* by forming the abstract sum $K_1 + K_2$, defining $\phi = \phi_i$ on K_i, and adding Z_1 to Z_2. Two continuous cycles are *homologous* if there exists a chain (K, ϕ, Z) such that $K \supset K_1$ and K_2, $\phi = \phi_i$ on K_i, and $\partial Z = Z_1 - Z_2$. The cycles of a fixed dimension divide up into homology classes. Two classes are *added* by adding representative elements. In this way we define the homology groups of R based on continuous cycles with local coefficients $\{G_x\}$. That they are topological invariants of R and the system $\{G_x\}$ is an immediate consequence of the definition.

If R is the space of a complex, these groups of R are isomorphic to those of K with the same local coefficients. The identity map ϕ_1 of K attaches to a chain Z of K the continuous chain (K, ϕ_1, Z) of R. This chain mapping commutes with ∂, and therefore induces homomorphisms of the groups of K into those of R. That these are isomorphisms follows from the lemma: If (K', ϕ, Z') is a chain with boundary of the form (K, ϕ_1, Z), then there is a chain Z_1 of K such that $\partial Z_1 = Z$, and the difference $(K', \phi, Z') + (K, \phi_1, -Z_1)$ bounds a continuous chain in R. The lemma is proved in the usual way by using the simplicial approximation theorem to construct a map of the product complex $K' \times I$ $(I = (0, 1))$ into R. The needed chain is found in $K' \times I$ with local coefficients in the induced system.

18. Čech cycles

The only difficulty in the way of extending local coefficients to Čech cycles is that of constructing a system of local groups in the nerve K of a finite open covering when such a system is given in R. It is clear that the former must be chosen so as to induce the given system in R under the natural map $R \rightarrow K$. If R is sufficiently complicated, it is possible to construct in R a local system

which is not induced by a local system in any nerve.[5] Therefore we are forced to restrict ourselves to a system $\{G_x\}$ induced in R by a system $\{G_x^0\}$ in a fixed nerve K^0. We then admit only those coverings which are refinements of K^0, and we use in them the local groups induced by their natural projections into K^0. With this modification, the definitions of the Čech homology and cohomology groups and their multiplications proceed as before.

If R is the space of a complex K, a local system in R is one in $K = K^0$. Using invariance under subdivision, one proves in the customary way (see [15], §9) that the groups of K and the Čech groups of R are isomorphic, and the isomorphisms preserve the multiplications. We are thus led to a proof that the homology theory (coef. $\{G_x\}$) of a complex K is a topological invariant of the space K with the local groups $\{G_x\}$.

19. Überdeckung

In a complex K choose a reference point o (preferably a vertex), and for each cell σ let α_σ be a path in K from o to a point $x(\sigma)$ in σ. If $\sigma' < \sigma$, let $\alpha_{\sigma\sigma'}$ be a path in the closure of σ from $x(\sigma)$ to $x(\sigma')$. The closed path $\alpha_\sigma\alpha_{\sigma\sigma'}\alpha_{\sigma'}^{-1}$ is abbreviated by $\gamma_{\sigma\sigma'}$. As elements of the fundamental group F of K (origin o), the γ's have the property

$$\gamma_{\sigma\sigma'}\gamma_{\sigma'\sigma''} = \gamma_{\sigma\sigma''} \qquad \text{for } \sigma'' < \sigma' < \sigma. \tag{19.1}$$

By means of the α's, we map isomorphically the chains of K with local coef. $\{G_x\}$ into ordinary chains with coefficients in G_o as follows. The transform $\bar{f}$ of the chain f with local coefficients is defined by $\bar{f}(\sigma) = \alpha_\sigma^{-1}(f(\sigma))$. By means of this isomorphism, we define operators $\bar{\partial}$, $\bar{\delta}$ for chains $\bar{f}$ by: $\bar{\partial}\bar{f} = \overline{\partial f}$, $\bar{\delta}\bar{f} = \overline{\delta f}$. From (9.3) we obtain

$$\begin{aligned} \bar{\partial}\bar{f}(\sigma') &= \textstyle\sum_\sigma [\sigma':\sigma]\gamma_{\sigma\sigma'}(\bar{f}(\sigma)), & \sigma' < \sigma, \\ \bar{\delta}\bar{f}(\sigma'') &= \textstyle\sum_\sigma [\sigma:\sigma'']\gamma_{\sigma''\sigma}^{-1}(\bar{f}(\sigma)), & \sigma < \sigma''. \end{aligned} \tag{19.2}$$

Thus the system of ordinary chains (coef. G_o) with the special operators $\bar{\partial}$, $\bar{\delta}$ are isomorphic to the system of chains with local coef. $\{G_x\}$ with the ordinary operators ∂, δ. They determine therefore isomorphic homology groups. It is a corollary that the homology groups determined by $\bar{\partial}$, $\bar{\delta}$ are independent of the choices of the α's.

The system $\bar{f}$, $\bar{\partial}$, $\bar{\delta}$ just described is called an *Überdeckung* of K by Reidemeister [10; 11]. An advantage of this approach is that it lends itself more readily to a computation of the homology groups. One may attempt to simplify the incidence matrices $[\sigma':\sigma]\gamma_{\sigma\sigma'}$, with elements in the group ring of F, by the usual methods of transforming bases and consolidation (see W. Franz [2]).

---

[5] This is the case if R is not locally simply-connected, and if $\{G_x\}$ is the system $\{G_x^L\}$ of local group rings of §8.

20. Zero and 1-dimensional groups

If the fundamental group F of K and the operations of F in G_o are given, then one may compute the 0 and 1-dimensional homology groups without further knowledge of K. Choose a basis $\alpha_1, \cdots, \alpha_h$ of F and a basis $r_1, \cdots, r_s$ for the relations in F (each r is a product of α's representing the unit). Construct a 2-complex K' consisting of one vertex o', one edge for each α_i (likewise denoted by α_i) with both end-points at o', and, for each r_i, a 2-cell E_i whose boundary is the product r_i of the α's. Clearly F is also the fundamental group of K'. The operations of F in G_o determine local coefficients in K' leading to 0 and 1-dimensional homology and cohomology groups which we shall prove isomorphic to those of K. By the duality of §12, it suffices to prove this for the homology groups.

Define a map ϕ of K' into K so that $\phi(o') = o$, $\phi(\alpha_i)$ represents α_i, and ϕ is continuous. It is readily seen that a map ψ of a complex K'' in K is homotopic to a map ψ' which, on the 1-dimensional part K_1'' of K'', can be expressed as a product $\phi\psi''$ of ϕ and a map ψ'' of K_1'' in K' which maps each vertex into o' and each edge into a product of α's. Thus every continuous 0 and 1-cycle is homologous to one of the form (K', ϕ, Z) (see §17). It is a further consequence that the 0-cycle Z bounds in K' if (K', ϕ, Z) bounds in K. This shows that the 0^{th} homology groups of K and K' are isomorphic. To prove the same for the 1-dimensional groups, we must show that a homology relation in K of the form $\partial(E, \psi', f(E) = g) = (K', \phi, Z)$, where E is a 2-cell and $\psi'(\partial E)$ is a product r of the α's, is a consequence of the relations in K'. Let E be regarded as a hemisphere of a 2-sphere S^2 and let E' be the other hemisphere. The map ψ'' of the equator in K' extends to a continuous map ψ'' of E' in K'; for the product r is expressible in terms of the r_i. The map ψ' of E and $\phi\psi''$ of E' define a map ψ' of S^2 in K, and thereby a 2-cycle $(S^2, \psi', f(E) = f(-E) = g)$. Thus $\partial(E', \psi', f(E') = g) = (K', \phi, Z)$, and ψ' factors into $\phi\psi''$.

Using the above results we may describe the 0-dimensional groups quite easily. Let G_o' be the subgroup of elements of G_o which are fixed under every automorphism $\alpha \in F$. *The* 0^{th} *cohomology group of* K *is isomorphic to the group* G_o'. Let G_o'' be the subgroup of elements of G_o expressible in the form $\sum_i(\alpha_i(g_i) - g_i)$ where $\alpha_i \in F$, $g_i \in G_o$. *The* 0^{th} *homology group of* K *is isomorphic to the difference group* $G_o - G_o''$.

It is worth noting that a continuous image of an n-sphere $(n > 1)$ in K determines a group of spherical n-cycles for any local coefficients; for the sphere is simply connected. However these cycles may or may not bound according to the structure of the system of local groups. Consider, as an example, the projective plane P^2 and the double covering of it by the 2-sphere S^2. With twisted integer coefficients (§14), the 2-cycles on P^2 form an infinite cyclic group and are nonbounding. The even multiples of the generator are images of the 2-cycles on S^2 with integer coefficients. Yet with simple integer coefficients in P^2, the mage of every 2-cycle on S^2 is bounding (algebraically zero).

THE UNIVERSITY OF MICHIGAN

Bibliography

1. P. Alexandroff-H. Hopf, *Topologie*, Berlin, 1936.
2. W. Franz, *Überdeckungen topologischer Komplexe mit hyperkomplexen Systemen*, J. reine angew. Math., 173 (1935), 174–184.
3. ———, *Über die Torsion einer Überdeckung*, ibid., 173 (1935), 245–254.
4. ———, *Torsionideale, Torsionklassen und Torsion*, ibid., 176 (1936), 113–124.
5. A. Komatu, *Über die Dualitätssätze der Überdeckungen*, Jap. Journ. Math., 13 (1937), 493–500.
6. ———, *Über die Überdeckungen von Zellenräumen*, Proc. Imp. Acad. Tokyo, 14 (1938), 340–341; 15 (1939), 4–7; 16 (1940), 55–58.
7. S. Lefschetz, *Topology*, Colloquium lectures, New York, 1930.
8. K. Reidemeister, *Homotopiegruppen von Komplexen*, Hamburger Abh., 10 (1934).
9. ———, *Homotopieringe und Linsenräume*, ibid., 11 (1935).
10. ———, *Überdeckungen von Komplexen*, J. reine angew. Math., 173 (1935), 164–173.
11. ———, *Topologie der Polyeder*, Leipzig, 1938.
12. ———, *Durchschnitt und Schnitt von Homotopieketten*, Monats. für Math., 48 (1939), 226–239.
13. G. de Rham, *Sur la théorie des intersections et les integrales multiples*, Comment. Math. Helv., 4 (1932), 151–157.
14. ———, *Sur les complexes avec automorphismes*, ibid., 12 (1940), 191–211.
15. N. Steenrod, *Universal homology groups*, Amer. Journ. Math., 58 (1936), 661-701.
16. ———, *Topological methods for the construction of tensor functions*, Annals of Math., 43 (1942).
17. H. Whitney, *On products in a complex*, Annals of Math., 39 (1938), 397–432.
18. ———, *On the theory of sphere-bundles*, Proc. Nat. Acad. Sci., 26 (1940), 148–153.
19. ———, *On the topology of differentiable manifolds*, Lectures in Topology, Univ. of Michigan, 1940.

ANNALS OF MATHEMATICS
Vol. 44, No. 4, October, 1943

NON-COMMUTATIVE CHAINS, II

BY WILLIAM W. FLEXNER[1]

(Received April 1, 1943)

Non-commutative chains of dimension two are here set up for a finite abstract system and by means of them an abelian "homology group" μ_2 is found in the spirit of the author's earlier definition of the Poincaré group as a homology group[2]. The group μ_2 is subdivision invariant and, furthermore, each finite simplicial complex determines a family of systems all having the same μ_2. At the end of this paper μ_2 is computed for some simplicial complexes including the lens spaces. For the lens spaces μ_2 is always the identity.

A recent paper by Hopf[3] has obvious points of contact with this one: each 2-syllable of this paper has as boundary one of the "loop paths" defining the "homotopy boundary" of the 2-cell in question in the sense of Hopf; if the C of (1.2) below is a cycle, the chain $E_{i_1}^{x_1} + \cdots + E_{i_s}^{x_s}$ is a Hopf spherical cycle.

1. The abstract *system* S consists of "cells", each having associated with it an integer called its dimension, and an operator F (meaning boundary) whose domain is the set of cells of S and whose range is a subset of the "chains" of S. The cells comprize the neutral cell, 1, and n-dimensional cells (n-cells, E_i^n) in finite number for $n = 0, 1, 2, 3$. It is convenient to suppose that 1 is an n-cell for each n.

By an n-chain when $n = 0, 1$ (but not 2, 3) is meant a class of "words" in the sense of the theory of non-abelian groups with a finite number of generators, subject to the relation $EE^{-1} = 1$, the letters of a word being n-cells or their inverses, for instance:

$$C = (E_{i_1}^n)^{x_1}(E_{i_2}^n)^{x_2} \cdots (E_{i_s}^n)^{x_s}, \qquad x_j = \pm 1. \tag{1.1}$$

The inverse of a chain and the product of two chains are then defined in the obvious way.[2] For the cell 1 and for zero and one-cells E, FE is defined as follows:

$$F1 = 1$$

$$FE_i^0 = 1 \text{ for zero-cells } E_i^0$$

$$FE_i^1 = E_{i_1}^0(E_{i_2}^0)^{-1} \text{ for one-cells } E^1 \text{ where } i_1, i_2 \text{ depend on } i.$$

As in N.C., $E_{i_1}^0$ and $(E_{i_2}^0)^{-1}$ are called the *outset* and *finish* of E_i^1. A word (1.1) has property ϕ if it $\equiv 1$ or if the finish of $(E_{i_j}^1)^{x_j}$ for $j = 1, 2, \cdots, s - 1$ is the

[1] Presented to the Society, December 1941.

[2] W. W. Flexner, Non-commutative Chains and the Poincaré Group, Duke Journal 8 (1941) pp. 497–505. Here referred to as N.C.

[3] Hopf, Heinz, *Fundamentalgruppe und zweite Bettische Gruppe*, Comment. Math. Helv. 14 (1942) pp. 257–309.

inverse of the outset of $(E^1_{i_{j+1}})^{x_{j+1}}$. The outset of $(E^1_{i_1})^{x_1}$ is called the outset of C^1 and the finish of $(E^1_{i_s})^{x_s}$ is called the finish of C^1. The system S is now subjected to condition Γ of N.C. p. 498 making S connected. If C^1 has property ϕ outset E^0 and finish $(E^0)^{-1}$, C is said to be a *one-cycle on* E^0. The one-cycles on E^0 form a subgroup $\mathfrak{Z}_1$ of the group of one-chains.

To each 2-cell E^2_i a one-cycle FE^2_i on some fixed zero-cell, call it ω, is now associated. (This FE^2_i bears to the boundary defined in N.C., p. 498, call it $\bar{F}E^2_i$, the relation $FE^2_i = a\bar{F}E^2_i a^{-1}$, where a is a 1-chain having property ϕ and outset ω). If a_i is a one-cycle on ω and E_i is a 2-cell or the inverse of a 2-cell the following symbolic combination Ω_i of a_i and E_i is called a *2-syllable*:

$$\Omega_i = a_i E_i a_i^{-1}.$$

The inverse and boundary of Ω_i are defined to be

$$\Omega_i^{-1} = a_i E_i^{-1} a_i^{-1}$$

$$F\Omega_i = a_i (FE_i) a_i^{-1}.$$

If C and D are "words" composed of 2-syllables, for instance,

$$(1.2)\qquad \begin{aligned} C &= (a_1 E^{x_1}_{i_1} a_1^{-1})(a_2 E^{x_2}_{i_2} a_2^{-1}) \cdots (a_s E^{x_s}_{i_s} a_s^{-1}), & x_k &= \pm 1,\\ D &= (b_1 E^{y_1}_{j_1} b_1^{-1})(b_2 E^{y_2}_{j_2} b_2^{-1}) \cdots (b_t E^{y_t}_{j_t} b_t^{-1}), & y_k &= \pm 1,\end{aligned}$$

let $CD = (a_1 E^{x_1}_{i_1} a_1^{-1}) \cdots (a_s E^{x_s}_{i_s} a_s^{-1})(b_1 E^{y_1}_{j_1} b_1^{-1}) \cdots (b_t E^{y_t}_{j_t} b_t^{-1})$ and

$$FC = (a_1 F E^{x_1}_{i_1} a_1^{-1}) \cdots (a_s F E^{x_s}_{i_s} a_s^{-1}).$$

Then

$$F(CD) = (FC)(FD).$$

If subjected to the following relations between the syllables these words form the group $\mathfrak{C}_2$ of two-chains:

$$(\text{r}.1)\qquad \Omega_i \Omega_i^{-1} = 1$$

$$(\text{r}.2)\qquad \Omega_i \Omega_j \Omega_i^{-1} = (F\Omega_i)\Omega_j(F\Omega_i)^{-1}$$

$$(\text{r}.3)\qquad a_i 1 a_i^{-1} = 1, \quad a_i \in \mathfrak{Z}_1 .$$

DEFINITION 1.3. *If b is a one-cycle on ω and C is defined by* (1.2), *then*

$$(1.4)\qquad b\,C\,b^{-1} = (b a_1 E^{x_1}_{i_1} a_1^{-1} b^{-1}) \cdots (b a_s E^{x_s}_{i_s} a_s^{-1} b^{-1}).$$

THEOREM 1.5. *If C is a 2-chain, b is a 1-cycle on ω and $C = 1$, then $b\,C\,b^{-1} = 1$.*

PROOF. The relations r which are used on the r.h.s. of (1.2) to reduce it to 1 can be used similarly on the r.h.s. of (1.4) and will reduce $b\,C\,b^{-1}$ to 1.

DEFINITION 1.6. *That two words C and D are identical as words will be written $C \equiv D$, that they define the same element of $\mathfrak{C}_2$ will be written $C = D$.*

THEOREM 1.7. *If $C = D$ then $FC = FD$.*

PROOF. The operator F applied to the l.h.s. of the relations r gives the same result as when applied to the r.h.s.

This shows that FC depends only on the element of $\mathcal{C}_2$ defined by C. Those elements $C \in \mathcal{C}_2$ for which $FC = 1$ form a subgroup $\mathcal{Z}_2$ of $\mathcal{C}_2$ called the group of two-cycles of S on ω.

THEOREM 1.8. *Each $C \in \mathcal{Z}_2$ commutes with every $D \in \mathcal{C}_2$.*

PROOF. By (r.2) $C\,D\,C^{-1}D^{-1} = (FC)\,D\,(FC)^{-1}D^{-1}$

$= D\,D^{-1}$ since C is a cycle

$= 1$ by (r.1).

COROLLARY 1.9. *The group $\mathcal{Z}_2$ is abelian.*

For each 3-cell $\mathcal{E}_i$ of S let $F\mathcal{E}_i$ be a unique element of $\mathcal{Z}_2$. The elements C of $\mathcal{Z}_2$ which can be defined by words all of whose syllables are of the form

$$a(F\mathcal{E}_i)^q a^{-1}, \qquad q = \pm 1, \qquad a \in \mathcal{Z}_1$$

make up a subgroup, $\mathcal{F}_2$, of $\mathcal{Z}_2$. The factor group $\mu_2 = \mathcal{Z}_2/\mathcal{F}_2$ is called the *second homology group of non-commutative chains of S on ω.*

2. If S is a system then $\bar{S}$ is a system and is called an *elementary subdivision of* S if $\bar{S}$ is related to S as follows. The elements of $\bar{S}$ are obtained from those of S by replacing a single p-cell of S by two p-cells Y and Z and a $(p - 1)$-cell X, $p = 1, 2, 3$. If E_1 is a cell of S such that $E_1 \neq E$ and FE_1 does not contain E, then FE_1 in $\bar{S}$ is the same as FE_1 in S. If in S

(2.1) $FE = GL$, G, L $(p - 1)$-chains for $p > 1$; G a vertex and L the inverse of a vertex for $p = 1$, then in $\bar{S}$

(2.2) $FX = FL$

(2.3) $FY = GX$

(2.4) $FZ = X^{-1}L$.

If FE_1 in S contains E, FE_1 in $\bar{S}$ is the same as FE_1 in S except that E is replaced by YZ.[4]

If S' is obtained from S by a finite sequence of elementary subdivisions, S' is called a *subdivision* of S.

THEOREM 2.5. *If $\bar{S}$ is an elementary subdivision of S, $\bar{S}$ and S have the same groups μ_2.*

COROLLARY 2.6. *If S' is a subdivision of S, S' and S have the same groups μ_2.*

In future all cycles will be "on ω" unless exception is specifically made and the following notational conventions will be adhered to:

Script capitals, except $\mathcal{C}$, $\mathcal{Z}$ and $\mathcal{F}$, for 3-cells;
the printed capitals E, J, X, Y, Z, for 2-cells and their inverses;
other printed capitals, except F, for 2-chains;
small latin letters for one-chains;
small greek letters, except π, τ and μ, for zero-cells and their inverses.

[4] For $p = 2$ this should be compared with N.C. p. 499, II. If in N.C. substitutions are made as shown below, the two definitions correspond: $FE \to C_2FE\,C_2^{-1}$, $FE' \to C_2FE'C_2^{-1}$, $FE'' \to FE''$. This shows that the distinction is a notational one.

3. Proof of Theorem 2.5. The groups of $\bar{S}$ will be distinguished by a bar. The proof that $\mu_2 \approx \bar{\mu}_2$ is then made in the three cases: $p = 1, p = 2, p = 3$. When $p = 1$ the one-cell e is replaced by the one-cells y and z and the vertex ξ. Formulas (2.1)–(2.4) become

$$Fe = \gamma\lambda, \qquad F\xi = F\lambda, \qquad Fy = \gamma\xi, \qquad Fz = \xi^{-1}\lambda.$$

If E is a 2-cell, FE in $\bar{S}$ is obtained by replacing e in FE of S by yz. Each 2-syllable of $\bar{\mathfrak{C}}_2$ is of the form $a\ Ja^{-1}$, $J^{\pm 1}$ a 2-cell of $\bar{S}$, $a \in \bar{\mathfrak{Z}}_1$. If a contains y or z non-trivially it contains both together as $(yz)^{\pm 1}$ for otherwise a cannot be a one-cycle. Hence the 2-chains of $\bar{S}$ become 2-chains of S when yz is everywhere replaced by e. Thus in this case there is an isomorphism $I\bar{\mathfrak{Z}}_2 = \mathfrak{Z}_2$ such that $I\bar{\mathfrak{F}}_2 = \mathfrak{F}_2$ and so $\mu_2 \approx \bar{\mu}_2$.

4. $p = 2$. E is replaced by Y, Z and x where if

$$FE = gm, \quad \text{then} \tag{4.1}$$

$$Fx = Fm \tag{4.2}$$

$$FY = gx \tag{4.3}$$

$$FZ = x^{-1}m. \tag{4.4}$$

Make a change of basis for the 2-chains of $\bar{\mathfrak{C}}_2$ such that

$$Y \rightarrow \bar{E}Z^{-1} \tag{4.5}$$

$$J \rightarrow J \qquad \text{if } J \neq Y \text{ is a 2-cell of } \bar{S},$$

where $\bar{E}$ is a new symbol playing the role of a 2-cell of $\bar{S}$ and such that

$$F\bar{E} = FE.$$

Lemma 4.6. *If $\bar{C} \in \bar{\mathfrak{Z}}_2$ and Y has been eliminated from $\bar{C}$ by* (4.5), *there is a $\bar{D}$ containing neither x nor Z and such that $\bar{D} = \bar{C}$.*

Proof 4.7. If the word $\bar{C}$ has a syllable $M = axbJb^{-1}x^{-1}a^{-1}$, where J is a 2-cell of $\bar{S}$, (4.4) gives that $x = m(FZ)^{-1}$. Hence

$$\begin{aligned} M &= amFZ^{-1}m^{-1}a^{-1}(ambJb^{-1}m^{-1}a^{-1})\ amFZm^{-1}a^{-1} \\ &= amZ^{-1}m^{-1}a^{-1}(ambJb^{-1}m^{-1}a^{-1})amZm^{-1}a^{-1} \qquad \text{by (r.2).} \end{aligned} \tag{4.8}$$

So $\bar{C}$ has equals containing no x. Among these there is at least one word having the fewest syllables containing Z. Let $\bar{D}$ be such a word and suppose s is the number of its syllables containing Z. If $s = 0$, lemma 4.6 is proved. If there is a syllable aZ^qa^{-1} in the word $\bar{D}$ $(q = \pm 1)$, there must be another syllable $b\ Z^{-q}b^{-1}$ the x of whose boundary cancels the x in the boundary of $a\ Z\ a^{-1}$ in $F\bar{D} = 1$. As there is no loss in assuming $q = 1$, $\bar{D}$ can be written:

$$\bar{D} = U(a\ Z\ a^{-1})V(b\ Z^{-1}b^{-1})W \qquad \text{where } U, V, W \text{ are elements of } \bar{\mathfrak{C}}_2. \tag{4.9}$$

Let $FU = u$, $FV = v$, $FW = w$. Then

$$F\bar{D} = u(a\,x^{-1}m\,a^{-1})v(b\,m^{-1}x\,b^{-1})w = 1,$$

and $a^{-1}v\,b = 1$ so $a = v\,b$. Therefore

$$(4.10)\qquad \begin{cases} \bar{D} = U(v\,b\,Z\,b^{-1}v^{-1})V(b\,Z^{-1}b^{-1})W \\ \quad = UV(b\,Z\,b^{-1})V^{-1}V(b\,Z^{-1}b^{-1})W \\ \quad = UVW, \end{cases}$$

which contradicts the hypothesis that $\bar{D}$ in its original form had s a minimum. Hence $s = 0$ and 4.6 is proved.

If $\bar{C} \in \bar{\mathfrak{Z}}_2$ and $\bar{C}$ as a word is free of x and Z, let $\tau\bar{C}$ be the chain of $\mathfrak{Z}_2$ obtained by replacing $\bar{E}$ in $\bar{C}$ by E in $\tau\bar{C}$. To see that $\tau\bar{C}$ depends only on the *element* $\bar{C}$ and not on the *word* $\bar{C}$ chosen to represent that element, observe that all steps (r.1), (r.2) and (r.3) which can be made for $\bar{C}$ can be reproduced for $\tau\bar{C}$ except those involving x and Z. But (4.8) and (4.9), (4.10) show that lemma 4.6 has the effect on the original word $\bar{C}$ of replacing x by m and then replacing Z by 1. Hence if m and 1 are used in $\tau\bar{C}$ where x and Z were used in $\bar{C}$ and if $F1$ be written $m^{-1}m$, the transformations of the word $\bar{C}$ can be duplicated throughout the word $\tau\bar{C}$.

Now if $C \in \mathfrak{Z}_2$, replacing E by $YZ = \bar{E}$ will yield a chain $\bar{C} \in \bar{\mathfrak{Z}}_2$ such that $\tau\bar{C} = C$, so $\tau\bar{\mathfrak{Z}}_2 = \mathfrak{Z}_2$. It is immediate that τ is a homomorphism. If $\mathfrak{H}$ is a 3-cell then $F\mathfrak{H}$ in $\bar{S}$ is by definition $F\mathfrak{H}$ in S with E replaced by $YZ = \bar{E}$. So $\tau F\mathfrak{H} = F\mathfrak{H}$ and $\tau^{-1}F\mathfrak{H} = F\mathfrak{H}$, whence $\tau^{-1}\mathfrak{F}_2 = \bar{\mathfrak{F}}_2$ and so[5] $\mu_2 \approx \bar{\mu}_2$ when $p = 2$.

5. $p = 3$. $\mathfrak{E}$ is replaced by $\mathfrak{Y}$, $\mathfrak{G}$ and X.

$$(5.1)\qquad F\mathfrak{E} = GL$$

$$(5.2)\qquad FX = FL$$

$$(5.3)\qquad F\mathfrak{Y} = GX$$

$$(5.4)\qquad F\mathfrak{G} = X^{-1}L.$$

If $\bar{C} \in \bar{\mathfrak{Z}}_2$ obtain $\tau\bar{C}$ by replacing X everywhere by L. By (5.2), $\tau\bar{C}$ is a unique element of $\mathfrak{Z}_2$. If $C \in \mathfrak{Z}_2$, then $C \in \bar{\mathfrak{Z}}_2$ and $\tau C = C$ so τ is a homomorphism such that $\tau\bar{\mathfrak{Z}}_2 = \mathfrak{Z}_2$. Since $\mathfrak{H} \neq \mathfrak{Y}$, $\mathfrak{G}$ a 3-cell of $\bar{S}$ implies $F\mathfrak{H}$ is the same in S and $\bar{S}$, $\tau(F\mathfrak{H}) = F\mathfrak{H}$. Also

$$\tau F\mathfrak{Y} = \tau(GX) = GL = F\mathfrak{E}$$

$$\tau F\mathfrak{G} = \tau(X^{-1}L) = 1 = F1,$$

so $\bar{C} \in \bar{\mathfrak{F}}_2$ implies $\tau\bar{C} \in \mathfrak{F}_2$, whence $\tau\bar{\mathfrak{F}}_2 \subset \mathfrak{F}_2$. It remains to show that $\tau^{-1}\mathfrak{F}_2 \subset \bar{\mathfrak{F}}_2$ from which $\tau^{-1}\mathfrak{F}_2 = \bar{\mathfrak{F}}_2$ and $\mu_2 \approx \bar{\mu}_2$ will follow.

[5] B. L. van der Waerden, *Moderne Algebra*, Berlin 1930, vol. 1, p. 136.

Adopt the following notation for 2-syllables of $\bar{S}$.

$\Omega_i = b_i J_i b_i^{-1}$, where $J_i \neq X, X^{-1}$ is a 2-cell of $\bar{S}$ or its inverse,

$\Lambda_i = a_i X^{q_i} a_i^{-1}$, where $q_i = \pm 1$,

$\Gamma_i = a_i L^{q_i} a_i^{-1}$. Then if

$$\bar{C} = \Omega_0 \Lambda_1 \Omega_1 \cdots \Lambda_s \Omega_s \tag{5.5}$$

$$\tau \bar{C} = \Omega_0 \Gamma_1 \Omega_1 \cdots \Gamma_s \Omega_s . \tag{5.6}$$

Take $C \in \mathfrak{F}_2$ and $\bar{C} \in \mathfrak{Z}_2$ such that $\tau\bar{C} = C$. Then C can be written in the form of the r.h.s. of (5.6) and $\bar{C}$ is given by (5.5). But (5.4) yields $X = L(F\mathcal{J})^{-1}$ so

$$\bar{C} = \Omega_0[a_1(LF\mathcal{J}^{-1})^{q_1}a_1^{-1}]\Omega_1 \cdots [a_s(LF\mathcal{J}^{-1})^{q_s}a_s^{-1}]\Omega_s . \tag{5.7}$$

If (5.7) be further expanded by definition 1.3, $\bar{C}$ becomes a product of syllables of the three kinds: Ω_i, Γ_i, $a_i F\mathcal{J}^{-q_i}a_i^{-1} = \Delta_i$. But N.C. p. 501 formula (3.3) now applies if

Δ_i be represented by the b_i of N.C.,

Ω_i be represented by a_i for some i,

Γ_i be represented by a_i for the remaining i.

Then (N.C.3.5) yields $C = g'$, and h_i will be a product of terms Ω_k and Γ_j. But with h_i occurs h_i^{-1} so by (r.2) each h_i may be replaced by a product of terms $F\Omega_k$ and $F\Gamma_j$. Hence each $h_i b_i h_i^{-1}$ of (N.C.3.4) is $h_i \Delta_m h_i^{-1} \in \bar{\mathfrak{F}}_2$ so

$$\bar{C} = D\, C \qquad \text{with } D \in \bar{\mathfrak{F}}_2 . \tag{5.8}$$

Since by hypothesis $C \in \mathfrak{F}_2$, C is a product of terms $c\mathcal{K}c^{-1}$ where $\mathcal{K}$ is a 3-cell or its inverse. If $\mathcal{K} \neq \mathcal{E}$, $F\mathcal{K}$ is the same in S and $\bar{S}$; if $\mathcal{K} = \mathcal{E}$, $F\mathcal{K} = F(\mathcal{Y}\mathcal{J})$ in $\bar{S}$, so in all cases $C \in \bar{\mathfrak{F}}_2$. Hence, by (5.8), $\bar{C} \in \bar{\mathfrak{F}}_2$ and $\tau^{-1}\mathfrak{F}_2 \in \bar{\mathfrak{F}}_2$, which completes the proof of theorem 2.5 and its corollary, 2.6.

6. Now suppose K is a finite simplicial connected 3-complex. It will be shown in this section that a system S can be associated with K, not in a unique manner, but so that the group μ_2 of S is uniquely determined by K. From 2.6 it then follows that μ_2 is unchanged by subdivision of K. Therefore for π_1 (see N.C.) and μ_2, S bears the same relation to K that an "abstract complex" bears to K for the ordinary homology groups.

Let the p-cells of S be the p-cells of K for $p = 0, 1, 2, 3$. Take ω to be some vertex of K. If a is a 1-cell, set $Fa = \beta\gamma^{-1}$ where β, γ are the end points of a in either order. If E is a 2-cell, let c, d, e each be a one-cell on the boundary of E (or the inverse of such a one-cell) such that cde is one of the six cyclic orders of the one-cells around E that has property ϕ. Let b be a one-chain with property ϕ outset ω and finish equal to the inverse of the outset of c. Then

$$FE = b(cde)b^{-1}.$$

If $\mathfrak{S}$ is a 3-cell, let

$$F\mathfrak{S} = (b_1 J_1^{q_1} b_1^{-1}) \cdots (b_4 J_4^{q_4} b_4^{-1}), \qquad q_i = \pm 1,$$

where the J_i are the four faces of $\mathfrak{S}$, b_i is a one-cycle and $FF\mathfrak{S} = 1$. That these conditions on $F\mathfrak{S}$ can be satisfied is easily proved by an example (see no. 13 below). They can in fact be satisfied in many different ways for each $\mathfrak{S}$.

A μ_2 being thus defined for K, it remains to show that μ_2 is independent of the particular application of the rules just given for finding S from K. This is the result of theorems 6.1, 6.5 and 8.1. (Theorem 8.1 depends on 2.6.)

THEOREM 6.1. *Let x be a one-chain of S having property ϕ, outset $\bar{\omega}$ and finish ω^{-1}. Then if $\bar{S}$ is obtained from S by replacing $FJ = j$ in S by $\bar{F}J = xjx^{-1}$ in $\bar{S}$ for each 2-cell J, and, for each 3-cell $\mathfrak{S}$, replacing*

$$F\mathfrak{S} = (a_1 J_1 a_1^{-1}) \cdots (a_s J_s a_s^{-1}) \quad \textit{in } S \textit{ by} \tag{6.2}$$

$$\bar{F}\mathfrak{S} = (b_1 J_1 b_1^{-1}) \cdots (b_s J_s b_s^{-1}) \quad \textit{in } \bar{S}, \textit{ where} \tag{6.3}$$

$$b_i = x\, a_i x^{-1},$$

then $\mu_2(S) \approx \mu_2(\bar{S})$.

This theorem shows that μ_2 does not depend on the particular point ω.

PROOF OF THEOREM 6.1. Use a bar to distinguish the groups and boundary operator of $\bar{S}$. To each syllable $\Omega_i = a_i J_i a_i^{-1}$ of S assign the syllable

$$\tau\Omega_i = b_i J_i b_i^{-1} \text{ of } \bar{S}.$$

Then

$$\bar{F}(\tau\Omega_i) = xF\Omega_i x^{-1}. \tag{6.4}$$

If $\tau(\Omega_1\Omega_2 \cdots \Omega_s)$ is set equal to $(\tau\Omega_1)(\tau\Omega_2) \cdots (\tau\Omega_s)$, (6.4) shows that $\tau\mathfrak{Z}_2 = \bar{\mathfrak{Z}}_2$. Since $x(F\Omega)ax^{-1} = (xF\Omega x^{-1})(xax^{-1})$, (r.2) and (6.4) show that

$$\begin{aligned} \tau(F\Omega_1\, \Omega_2\, F\Omega_1^{-1}) &= (xF\Omega_1 x^{-1})\tau\Omega_2(xF\Omega_1^{-1}x^{-1}) \\ &= (\bar{F}\tau\Omega_1)\tau\Omega_2(\bar{F}\tau\Omega_1)^{-1} = \tau(\Omega_1\Omega_2\Omega_1^{-1}), \end{aligned}$$

so τC, $C \in \mathfrak{Z}_2$, is a unique element of $\bar{\mathfrak{Z}}_2$, and τ is a homorphism of $\mathfrak{Z}_2$ into $\bar{\mathfrak{Z}}_2$. Now by (6.2) and (6.3) $\tau F\mathfrak{S} = \bar{F}\mathfrak{S}$ and $\tau^{-1}\bar{F}\mathfrak{S} = F\mathfrak{S}$, whence $\mu_2 \approx \bar{\mu}_2$.

THEOREM 6.5. *If $\bar{S}$ is obtained from S by replacing, for a single 2-cell E of S, $FE = e$ by $\bar{F}E = xex^{-1}$ where $x \in \mathfrak{Z}_1$, and replacing E where it occurs in the boundary $F\mathfrak{S}$ of three-cells $\mathfrak{S}$ by $x^{-1}Ex$ in $\bar{F}\mathfrak{S}$, then $\mu_2(S) \approx \mu_2(\bar{S})$.*

Since replacing e by e^{-1} is a trivial change, this theorem shows that μ_2 is independent of the particular choices of the boundaries of the 2-cells. For instance if $FE = a(rst)a^{-1}$ and $\bar{F}E = b(str)b^{-1}$, putting $x = br^{-1}a^{-1}$ brings the change under 6.5.

PROOF OF THEOREM 6.5. In chains of $\bar{S}$ replace E wherever it occurs by $x\bar{E}x^{-1}$, where $\bar{E}$ is a new symbol such that $\bar{F}\bar{E} = e$. This trivial change of basis for the 2-chains of $\bar{S}$ makes the cycles and boundaries of $\bar{S}$ formally identical with those of S.

7. Theorem 8.1 below shows that μ_2 does not depend on the particular choice of the boundaries of the 3-cells made in determining S from K. The following purely algebraic lemmas are used in its proof.

LEMMA 7.1. *If in a free multiplicative group the element $E \neq 1$ contains, in normal form,* (N.C.pp. 497–8), *no repeated letter, then*

$$D^{-1}ED = E \tag{7.2}$$

implies that the element D is a power of E.

If $D = 1$ the lemma is trivial, so assume that $D \neq 1$, and suppose that D is given in normal form by

$$D = b_1 \cdots b_p c_{p+1} \cdots c_q b_p^{-1} \cdots b_1^{-1} \tag{7.3}$$

where $q > p$ and $c_{p+1}c_q \neq 1$. Suppose further that, in normal form,

$$E = a_1 a_2 \cdots a_n \tag{7.4}$$

where

$$a_i = a_j^{\pm 1} \quad \text{implies} \quad i = j. \tag{7.5}$$

Then

$$b_1 \cdots b_p c_q^{-1} \cdots c_{p+1}^{-1} b_p^{-1} \cdots b_1^{-1} a_1 \cdots a_n b_1 \cdots b_p c_{p+1} \cdots c_q b_p^{-1} \cdots b_1^{-1} = a_1 \cdots a_n . \tag{7.6}$$

Hence either, case 1, $a_1 = b_1$, or, case 2, $a_n^{-1} = b_1$, since the l.h.s. of (7.6) must be reducible to the r.h.s. which is normal.

Case 1. If $a_1 = b_1$, (7.6) yields

$$b_2 \cdots b_p c_q^{-1} \cdots c_{p+1}^{-1} b_p^{-1} \cdots b_2^{-1} a_2 \cdots a_n a_1 b_2 \cdots b_p c_{p+1} \cdots c_q b_p^{-1} \cdots b_2^{-1} = a_2 \cdots a_n a_1 \tag{7.7}$$

and for the next step in the reduction of the l.h.s. it is necessary that either $a_2 = b_2$ or $a_1^{-1} = b_2$. But if $a_1^{-1} = b_2$, then $b_1 = b_2^{-1}$ which contradicts the normality of (7.3). Hence $a_2 = b_2$ and another step similar to that from (7.6) to (7.7) is possible and so on. After p steps there result

$$b_i = a_j, \qquad 1 \leqq i \leqq p, \qquad i \equiv j \bmod n, \tag{7.8}$$

and

$$c_q^{-1} \cdots c_{p+1}^{-1} a_{r+1} \cdots a_n a_1 \cdots a_r c_{p+1} \cdots c_q = a_{r+1} \cdots a_n a_1 \cdots a_r \tag{7.9}$$

where

$$p = sn + r, \qquad s \text{ integral}, \qquad 0 \leqq r < n. \tag{7.10}$$

Again normality of (7.3) gives $a_r \neq c_{p+1}^{-1}$, so

$$c_{p+1} = a_{r+1} \tag{7.11}$$

and

$$(7.11.1)\qquad c_{p+1}^{-1} c_q^{-1} \cdots c_{p+2}^{-1} a_{r+2} \cdots a_n a_1 \cdots a_r a_{r+1} c_{p+2} \cdots c_p c_{p+1} = a_{r+2} \cdots a_n a_1 \cdots a_{r+1}.$$

Then $c_{p+2} = a_{r+2}$ and a further step is possible. After $q - p$ such steps

$$(7.12)\qquad c_i = a_j, \qquad p < i \leqq q, \qquad i \equiv j \bmod n,$$

and

$$(7.13)\qquad c_q^{-1} \cdots c_{p+1}^{-1} a_{m+1} \cdots a_n a_1 \cdots a_m c_{p+1} \cdots c_q = a_{m+1} \cdots a_n a_1 \cdots a_m$$

where

$$(7.14)\qquad q = kn + m, \qquad k \text{ integral}, \qquad 0 \leqq m < n.$$

So by (7.14) $c_{p+1} = a_{m+1}$. But by (7.11) $c_{p+1} = a_{r+1}$ so, using (7.5) $m = r$ and by (7.10) and (7.14)

$$(7.15)\qquad q - p \text{ is a multiple of } n.$$

Using (7.8) and (7.12) with (7.15)

$$(7.16)\qquad b_p = a_r = a_m = c_q.$$

But if $p \neq 0$, (7.16) contradicts the normality of (7.3), so $p = 0$ and by (7.15) q is a multiple of n. This with (7.12) proves the lemma in case 1.

CASE 2. Replacing (7.2) by $D^{-1}E^{-1}D = E^{-1}$ and renaming the letters of the word E so that $E = a_n^{-1}a_{n-1}^{-1} \cdots a_1^{-1}$ makes case 2 formally the same as case 1 and so provides, in case 2, a proof of the lemma for E^{-1} and hence for E.

LEMMA 7.17. *Lemma 7.1 holds if (7.2) is replaced by $D^{-1}E\,D = E^{-1}$.*

The proof is just as before except that in the formulas whose numbers follow the right hand side must be replaced by its inverse: (7.6), (7.7), (7.11.1), (7.13).

COROLLARY 7.18. *If E is as in lemma 7.0, $D^{-1}E\,D = E^{-1}$ implies $E = 1$.*

8. THEOREM 8.1. *If $\bar{S}$ is obtained from S by replacing*

$$F\mathfrak{E} = (b_1J_1b_1^{-1}) \cdots (b_4J_4b_4^{-1}) \text{ in } S,$$

where $\mathfrak{E}$ is a 3-cell of S, J_i is a 2-cell of S or the inverse of such a 2-cell by

$$(8.2)\qquad \bar{F}\mathfrak{E} = (c_1J_{k_1}^{r_1}c_1^{-1}) \cdots (c_4J_{k_4}^{r_4}c_4^{-1}),$$

where k_1, k_2, k_3, k_4 is a permutation of 1, 2, 3, 4 and $r_i = \pm 1$, then $\mu_2 = \bar{\mu}_2$.

PROOF. Here the groups $\mathfrak{Z}_2$ and $\bar{\mathfrak{Z}}_2$ are identical so the proof can be made by showing that $\mathfrak{F}_2 \equiv \bar{\mathfrak{F}}_2$. For this it is sufficient to show that $F\mathfrak{E} \in \bar{\mathfrak{F}}_2$ and $\bar{F}\mathfrak{E} \in \mathfrak{F}_2$, which can be done by finding one-cycles p and q such that

$$(8.3)\qquad (p\,F\mathfrak{E}p^{-1})(q\,\bar{F}\mathfrak{E}^{\pm 1}q^{-1}) = 1,$$

for then

$$F\mathfrak{E} = p^{-1}q\bar{F}\mathfrak{E}^{\mp 1}q^{-1}p \in \bar{\mathfrak{F}}_2$$

and

$$\bar{F}\mathfrak{E} = q^{-1}pF\mathfrak{E}^{\mp 1}p^{-1}q \in \mathfrak{F}_2.$$

Suppose that $r_1 = 1$. (If it were -1 the notation could be changed by replacing $\bar{F}\mathfrak{S}$ by its inverse). Let

$$U = (b_2J_2b_2^{-1})(b_3J_3b_3^{-1})(b_4J_4b_4^{-1}).$$

and write $FU = u$. Take a such that $k_a = 1$ in the permutation k_1, k_2, k_3, k_4.

Let
$$V = (c_1J_{k_1}^{r_1}c_1^{-1}) \cdots (c_{a-1}J_{k_{a-1}}^{r_{a-1}}c_{a-1}^{-1}) \qquad \text{if } a > 1$$
$$V = 1 \qquad \text{if } a = 1.$$

Let
$$W = (c_{a+1}J_{k_{a+1}}^{r_{a+1}}c_{a+1}^{-1}) \cdots (c_4J_{k_4}^{r_4}c_4^{-1}) \qquad \text{if } a < 4$$
$$W = 1 \qquad \text{if } a = 4.$$

Write $FV = v$ and $FW = w$. Then

$$F\mathfrak{S} = b_1J_1b_1^{-1}U$$

$$\bar{F}\mathfrak{S} = Vc_1J_1c_1^{-1}W = Vc_1J_1c_1^{-1}V^{-1}VW = vc_1J_1c_1^{-1}v^{-1}VW.$$

Let

$$G = (b_1^{-1}F\mathfrak{S}^{-1}b_1)(c_1^{-1}v^{-1}\bar{F}\mathfrak{S}vc_1). \tag{8.4}$$

Comparing (8.4) and (8.3) shows that a proof that $G = 1$ will prove theorem 8.1. But $G = (b_1^{-1}U^{-1}b_1)(c_1^{-1}v^{-1}VWvc_1)$ contains no 2-cells except J_2, J_3 and J_4. Hence G is a chain of the subsystem $\hat{S}$ of S containing the zero and one-cells of S and J_2, J_3, J_4 but no other 2-cells and no 3-cells at all. By the inverse of subdivision, using theorems 6.1 and 6.5 as often as necessary, $\hat{S}$ can be replaced by S' in which J_2, J_3, J_4 are united to form a single 2-cell E. Theorems 2.5, 6.1, 6.5 show that $\mu_2(\hat{S}) \approx \mu_2(S')$. So proving that $\mu_2(S') \approx 1$ will show that $\mu_2(\hat{S}) \approx 1$ and hence that the G of (8.4) considered as a chain of $\hat{S}$ is in $\mathfrak{F}_2$. But $\hat{S}$ has no 3-cells so this is tantamount to proving that $G = 1$ in $\hat{S}$. Finally, since $\hat{S}$ is a subsystem of S, $G = 1$ in $\hat{S}$ implies $G = 1$ in S. So everything hinges on showing $\mu_2(S') \approx 1$, which is the result of the following theorem.

9. Theorem 9.1. *If S is a system having a single 2-cell E and no 3-cells, then $\mu_2(S) \approx 1$.*

Proof. By theorem 6.1 there is no loss of generality in assuming that ω is the outset of FE. Let $FE = e = xyz$, where x, y, z are one-cells of S or their inverses. Since E is a simplex, x, y and z are distinct. Theorem 9.1 is equivalent to the following lemma.

Lemma 9.2. *If*

$$D \equiv (a_1E^{q_1}a_1^{-1}) \cdots (a_sE^{q_s}a_s^{-1}) \qquad \textit{where } q_i = \pm 1, \tag{9.3}$$

defines an element of $\mathfrak{Z}_2$ for the S now under consideration, then $D = 1$.

Proof. Suppose the a_i of (9.3) are in normal form. If

$$a_i = be^rc, \qquad r = \pm 1, \tag{9.4}$$

then

$$(a_iE^{q_i}a_i^{-1}) = (bE^rb^{-1})(bcE^{q_i}c^{-1}b^{-1})(bE^{-r}b^{-1}),$$

so it may be assumed that no a_i in (9.3) can be written in the form (9.4). Now since x and its inverse must occur equally often in (9.3) (because $FD = 1$) s must be even. If $s = 2$, $FD = (a_1e^{q_1}a_1^{-1})(a_2e^{q_2}a_2^{-1}) = 1$ so $a_2^{-1}a_1e^{q_1}a_1^{-1}a_2 = e^{-q_2}$ and, by either lemma 7.1 or 7.17, $a_2^{-1}a_1 = e^t$ and $q_1 = -q_2$. Therefore $D = (a_2E^ta_2^{-1})(a_2E^{q_1}a_2^{-1})(a_2E^{-t}a_2^{-1})(a_2E^{-q_1}a_2^{-1}) = 1$ which proves lemma 9.2 for $s = 2$.

As hypothesis of an induction assume that S is such that all 2-cycles of S (having no a_i in form (9.4)) which can be written in less than s syllables define the identity of $\mathfrak{Z}_2$, i.e. if H is a cycle of less than s syllables, $H = 1$. This hypothesis has just been proved for $s = 3$. The D of (9.3) represents a cycle of s syllables. Suppose one of these syllables, Ω, contains a one-cell m which is not $x^{\pm 1}$, $y^{\pm 1}$ or $z^{\pm 1}$:

(9.4.1) $$\Omega = bmcE^qc^{-1}m^{-1}b^{-1} \qquad \text{where } \mathfrak{c} \text{ contains no } m.$$

Since $ce^qc^{-1} = 1$ implies $c = 1$ which is impossible, the m in Ω must, in $FD = 1$, cancel the m in some other syllable of D, i.e. for some Ω

(9.5) $$D \equiv M(bmcE^qc^{-1}m^{-1}b^{-1})\ N\ (dmgE^rg^{-1}m^{-1}d^{-1})\ T$$

where M, N, T are 2-chains, $r = \pm 1$ and

(9.6) $$b^{-1}nd = 1 \qquad \text{when } FN = n,$$

for otherwise $FD \neq 1$. Using (9.6) in (9.5)

(9.7) $$D = M[bm(cE^qc^{-1})(gE^rg^{-1})m^{-1}b^{-1}]NT.$$

If the boundary of the chain in the square brackets is 1, $D = 1$ by hypothesis of the induction. If not, the bracket can itself play the role of the expression (9.4.1) and the steps (9.5)–(9.7) can be repeated for the same m until either the hypothesis of the induction can be used or $D = pm^tD'm^{-t}p^{-1}$, where $D' = (h_1E^{r_1}h_1^{-1}) \cdots (h_sE^{r_s}h_s^{-1})$, $r_i = \pm 1$, $FD' = 1$. Though D' is not necessarily a chain, it is convenient to introduce it as a symbol (see definition 1.7) and to define FD' in the obvious way. Now D' contains m fewer times than D, all h_i have property ϕ, a common outset and the finish ω^{-1}.

If D' has a letter m' which is not $x^{\pm 1}$, $y^{\pm 1}$ or $z^{\pm 1}$, the same process may be applied to m' in D' as was to m in D. Since $pmD'm^{-1}p^{-1}$ is a cycle of S the same use of the hypothesis of induction can be made at the new steps as was made before. Eventually either lemma 9.1 is verified or $D = u\bar{D}u^{-1}$, where $\bar{D} = (k_1E^{p_1}k_1^{-1}) \cdots (k_sE^{p_s}k_s^{-1})$ and $p_i = \pm 1$ and all k_i have property ϕ, a common outset σ and finish ω^{-1} and are made up of the letters x, y z. Like D', $\bar{D}$ is not necessarily a chain.

Suppose σ is the outset of y. Then xk_i is a cycle on ω and so is a one-cycle of S and must, in normal form, be $(xyz)^{a_i} = e^{a_i}$, a_i an integer. Therefore $x\bar{D}x^{-1}$ is a chain such that

$$x\bar{D}x^{-1} = (e^{a_1}E^{p_1}e^{-a_1}) \cdots (e^{a_s}E^{p_s}e^{-a_s}) = E^{a_1}E^{p_1}E^{-a_1} \cdots E^{a_s}E^{p_s}E^{-a_s}$$

is a cycle of S and so $F(x\bar{D}x^{-1}) = 1$ which implies $p_1 + p_2 + p_3 + \cdots + p_s = 0$ which in turn implies $x\bar{D}x^{-1} = 1$. But $D = ux^{-1}(x\bar{D}x^{-1})xu^{-1} = ux^{-1}(1)xu^{-1} = 1$ by (r.3). If σ were the outset of z, similar reasoning would give $xy\bar{D}y^{-1}x^{-1} = 1$ and $D = 1$ again. If σ were ω, then $\bar{D} = 1$ directly and so $D = 1$. Now the hypothesis of the induction has been proved for s on the assumption that it holds for $s - 1$, so the proof of lemma 9.1 and hence of theorem 8.1 is complete.

10. In this and subsequent sections the theory just developed will be applied to some particular simplicial complexes. By corollary 2.6, the n-sphere S^n can be considered to be composed of two n-simplexes, $Y^n = \{\alpha_0\alpha_1 \cdots \alpha_n\}$ and $Z^n = \{\beta_0\beta_1 \cdots \beta_n\}$ where the identifications of lower dimensional simplexes given by $\{\alpha_{i_0}\alpha_{i_1} \cdots \alpha_{i_p}\} = \{\beta_{i_0}\beta_{i_1} \cdots \beta_{i_p}\}^{-1}$, $p < n$, have been made. So, n for > 2, the two-cycles of S^n are the two-cycles of the n-simplex Y^n and therefore $\mu_2(Y^n) \approx 1$ implies $\mu_2(S^n) \approx 1$. But for $n > 2$, $\mu_2(S^n) \approx 1$ implies $\mu_2(Y^{n+1}) \approx 1$. It follows that $\mu_2(Y^3) \approx 1$ implies $\mu_2(S^n) \approx 1$ for $n > 2$.

Now S^2 and $Y^3 = \mathcal{Y}$ can be treated together. Adopt the notation:

$$e_0 = \{\alpha_1\alpha_2\} = \{\beta_1\beta_2\}^{-1}, \qquad e_1 = \{\alpha_0\alpha_2\} = \{\beta_0\beta_2\}^{-1}, \qquad e_2 = \{\alpha_0\alpha_1\} = \{\beta_0\beta_1\}^{-1},$$

$$\alpha_i = \{\alpha_i\} = \{\beta_i\}^{-1}, \qquad \omega = \alpha_0,$$

$$Fe_0 = \alpha_1\alpha_2^{-1}, \qquad Fe_1 = \alpha_0\alpha_2^{-1}, \qquad Fe_2 = \alpha_0\alpha_1^{-1},$$

$$FY = e_2e_0e_1^{-1}, \qquad FZ = e_1e_0^{-1}e_2^{-1}, \qquad F\mathcal{Y} = YZ.$$

By theorem 2.5, Y and Z and e_1 can be replaced by E where $FE = 1$ and $F\mathcal{Y} = E$. Then e_2 and e_0 can be replaced by e where $Fe = \alpha_0\alpha_2^{-1}$. So it follows that the cycles of S^2 and $\mathcal{Y}$ are E^t, and so $\mu_2(S^2)$ is the infinite cyclic group whereas $\mu_2(\mathcal{Y}) \approx 1$. It follows that $\mu_2(S^n) \approx 1$ for $n > 2$. Since 1 is the only 2-cell of S^1 and S^0 it follows that $\mu_2(S^n) \approx 1$ for $n < 2$.

11. Let M be the complex composed of S^2 and an S^1 (called c) having the single vertex α_0 in common with S^2 (see no. 10). $Fc = \alpha_0\alpha_0^{-1} = 1$. Then $\mu_2(M)$ is the free abelian group with the infinite set c^tEc^{-t}, t integral, of generators. (Notice that $(c^tEc^{-t})(c^sE^{-1}c^{-s}) = 1$ implies $t = s$.)

Let N be a complex like M except that c bounds a 2-cell X: $FX = c$. Then $c^tEc^{-t} = X^tEX^{-t} = X^tX^{-t}E$ by theorem 1.8, whence $c^tEc^{-t} = E$. Therefore $\mu_2(N)$ is the infinite cyclic group.

12. Let K be the torus with vertex ω, one-cells a and b, and 2-cell E, where $Fa = Fb = \omega\omega^{-1}$, $FE = a^{-1}b^{-1}ab$. The only 2-cycle is the identity so μ_2 is the identity for the torus.

13. Lens spaces.[6] Take p 3-simplexes $\mathfrak{G}_0, \mathfrak{G}_1, \cdots, \mathfrak{G}_{p-1}$ such that $\mathfrak{G}_i$

[6] H. Seifert and W. Threlfall, *Lehrbuch der Topologie*, Berlin 1934, p. 210.

has on its boundary the 2-simplexes A_i, B_i, C_i, D_i and the one-simplexes a_i, b_i, c_i, d_i, e_i, f_i with boundary relations:

$$F\mathfrak{G}_i = A_iB_iD_iC_i, \qquad FA_i = a_ic_i^{-1}d_i, \qquad FB_i = d_i^{-1}b_ie_i,$$
$$FC_i = e_i^{-1}f_ia_i^{-1}, \qquad FD_i = e_i^{-1}(b_i^{-1}c_if_i^{-1})e_i.$$

Group the tetrahedra radially about the one-simplex $b = b_0 = b_1 = \cdots = b_{p-1}$ so that

$$(13.1) \qquad f_i = e_{i+1}, \qquad c_i = d_{i+1}, \qquad D_i = e_i^{-1}f_iB_{i+1}^{-1}f_i^{-1}e_i$$

if subscripts are reduced mod p. Thus grouped the tetrahedra form a "lens". The simplexes C_i compose one face of this lens, the simplexes A_i the other. If p and q are relatively prime and $q < p/2$, subject one lens face to a rotation of $2\pi q/p$ and identify it with the other. This amounts to letting

$$(13.2) \qquad a_i = a_{i+q}, \qquad e_i = d_{i+q}, \qquad f_i = c_{i+q}, \qquad C_i = A_{i+q}^{-1}.$$

Since the numbers kq, $k = 0, 1, \cdots, p-1$, are distinct mod p, the effect of (13.1) and (13.2) on the one-simplexes can be summed up in

$$a_i = a, \qquad b_i = b, \qquad e_i = f_{i-1} = c_{i+q-1} = d_{i+q} = i.$$

The number i on the extreme right is to be understood as an abbreviation for r_i where r is a fixed letter and i is a variable subscript. Subscripts are to be reduced mod p.

Then for the lens space (p, q),

$$(13.3) \qquad \begin{cases} F\mathfrak{G}_i = A_iB_i[i^{-1}(i+1)B_{i+1}^{-1}(i+1)^{-1}i]A_{i+q}^{-1} \\ FA_i = a(i-q+1)^{-1}(i-q) \\ FB_i = (i-q)^{-1}bi. \end{cases}$$

The system S determined by the equations (13.3) (plus easily deduced boundary relations for the one-cells, here omitted) is now simplified by repeated use of theorem 2.5. First let

$$(13.4) \qquad X_i = i^{-1}(i+1)B_{i+1}^{-1}(i+1)^{-1}i.$$

I. Replace $\mathfrak{G}_0$, $\mathfrak{G}_q$ and A_q by $\mathfrak{S}_1$ where

$$F\mathfrak{S}_1 = F(\mathfrak{G}_0\mathfrak{G}_q) = A_0B_0X_0B_qX_qA_{2q}^{-1}.$$

Replace $\mathfrak{S}_1$, $\mathfrak{G}_{2q}$ and A_{2q} by $\mathfrak{S}_2$ where

$$F\mathfrak{S}_2 = F(\mathfrak{S}_1\mathfrak{G}_{2q}) = A_0B_0X_0B_qX_qB_{2q}X_{2q}A_{3q}^{-1},$$

and so on up to and including the replacement of $\mathfrak{S}_{p-3}$, $\mathfrak{G}_{(p-2)q}$ and $A_{(p-2)q}$ by $\mathfrak{S}_{p-2}$ where

$$F\mathfrak{S}_{p-2} = F(\mathfrak{S}_{p-3}\mathfrak{G}_{(p-2)q}) = A_0(B_0X_0 \cdots B_{(p-2)q}X_{(p-2)q})A_{(p-1)q}^{-1}.$$

Replace A_0, $A_{(p-1)q}$ and a by J where $FJ = F(A_{(p-1)q}^{-1}A_0)$ and

$$F\mathfrak{E}_{p-2} = A_0[(B_0X_0 \cdots B_{(p-2)q}X_{(p-2)q})A^{-1}_{(p-1)q}A_0]A_0^{-1}$$
$$= (B_0X_0 \cdots B_{(p-2)q}X_{(p-2)q})J$$

by theorem 1.8 since the square bracket is a cycle; and, similarly,

$$F\mathfrak{G}_{(p-1)q} = J^{-1}B_{(p-1)q}X_{(p-1)q}.$$

Replace $\mathfrak{E}_{p-2}$, $\mathfrak{G}_{(p-1)q}$ and J by $\mathfrak{E}$ where

(13.5) $$F\mathfrak{E} = B_0X_0B_qX_q \cdots B_{(p-1)q}X_{(p-1)q}.$$

There now remain of the 3-cells only $\mathfrak{E}$, of the 2-cells only the B_i, and of the one-cells b and i, $i = 0, 1, \cdots, (p-1)$.

II. Now it is possible to eliminate all but one 2-cell and all the one-cells but b and $[(p-1)q]$, which last will also be written $(-q)$. Let

(13.6) $$Z_{kq} = (-q)^{-1}b^{1+k-p}(1+kq)B^{-1}_{1+kq}(1+kq)^{-1}b^{p-k-1}(-q).$$

Since it is an immediate consequence of (13.3) that

(13.7) $$F(B_{(k+1)q} \cdots B_{(p-1)q}) = (kq)^{-1}b^{p-k-1}(-q)$$

using (13.7) on (13.6) and then (r.2) on the r.h.s. of the result gives, by (13.4),

(13.8) $$X_{kq} = B_{(k+1)q} \cdots B_{(p-1)q}Z_{kq}B^{-1}_{(p-1)q} \cdots B^{-1}_{(k+1)q}.$$

Now using (13.8) to eliminate the X's from (13.5) gives

(13.9) $$F\mathfrak{E} = B_0B_q \cdots B_{(p-1)q}Z_0Z_q \cdots Z_{(p-1)q}.$$

Let D_{rq} be defined by

(13.10) $$D_{rq} = B_1B_{1+q} \cdots B_{1+rq}.$$

LEMMA. $Z_0Z_q \cdots Z_{rq} = (-q)^{-1}b^{1+r-p}(1+rq)D^{-1}_{rq}(1+rq)^{-1}b^{p-r-1}(-q)$.

Proof by induction on r. For $r = 0$, this is (13.6), so assume the lemma for $r < k$. Let

$$S = (Z_0 \cdots Z_{(k-1)q})Z_{kq} = (-q)^{-1}\{[b^{k-p}(1+(k-1)q)D^{-1}_{(k-1)q}(1+(k-1)q)^{-1}b^{p-k}]$$
$$\cdot[b^{1+k-p}(1+kq)B^{-1}_{1+kq}(1+kq)^{-1}b^{p-k-1}]\}(-q).$$

By (13.3), $(1+(k-1)q) = b(1+kq)(FB_{1+kq})^{-1}$ so

$$S = (-q)^{-1}b^{k-p+1}(1+kq)D^{-1}_{kq}(1+kq)^{-1}b^{p-k-1}(-q),$$

which proves the lemma.

It is a consequence of the lemma that

(13.11) $$Z_0Z_q \cdots Z_{(p-1)q} = (-q)^{-1}(1+(p-1)q)D^{-1}_{(p-1)q}(1+(p-1)q)^{-1}(-q).$$

(13.12) If $$1 + nq \equiv 0 \bmod p$$

then

$$1 + (p-1)q \equiv (p-n-1)q \bmod p$$

and

$$1 + (n - k)q \equiv (p - k)q \bmod p$$

for every k, so, by (13.10)

(13.13) $D_{(p-1)q} = B_1 \cdots B_{1+(n-1)q} B_0 B_q \cdots B_{(p-n-1)q} B_{(p-n)q} \cdots$
$$B_{(p-1)q} B^{-1}_{1+(n-1)q} \cdots B_1^{-1}.$$

But $F(B_1 \cdots B_{1+(n-1)q}) = (1-q)^{-1} b^n (1 + (n-1)q)$, so (13.13) becomes

(13.14) $D_{(p-1)q} = (1-q)^{-1} b^n (1 + (n-1)q) B_0 B_q \cdots$
$$B_{(p-1)q} (1 + (n-1)q)^{-1} b^{-n} (1-q).$$

Now combining (13.14), (13.11) and (13.9) and noticing that $(1 - q) \equiv (1 + (p-1)q) \bmod p$ yields

(13.15) $F\mathfrak{S} = B_0 B_q \cdots B_{(p-1)q}[(-q)^{-1} b^n (-q)(B_0 B_q \cdots B_{(p-1)q})^{-1} (-q)^{-1} b^{-n} (-q)]$.

The elimination of cells now takes place in $(p-1)$ steps. Let $G_0 = B_0$ and at the k^{th} step replace G_{k-1} and B_{kq} and $(kq - q)$ by G_k, where $FG_k = F(G_{k-1}B_{kq})$. Then letting $G_{(p-1)q} = G$ and $(-q)^{-1}b(-q) = y$, (13.15) becomes

(13.16) $$F\mathfrak{S} = G(y^n G^{-1} y^{-n}),$$

and, from the definition of G

(13.17) $$FG = y^p.$$

From (13.17) it follows that π_1 is the cyclic group of order p. Notice that by (r.2) and (13.17)

$$(y^{r+p} G^t y^{-r-p}) = y^r G^t y^{-r}.$$

Any 2-cycle, H, is a combination of syllables of the type $(y^r G y^{-r})$, and if $H \neq 1$, somewhere in H there must be an adjacent pair

$$M = (y^s G y^{-s})(y^t G^{-1} y^{-t})$$

or the inverse of such a pair, $s \not\equiv t \bmod p$, so that $H = LMN$. But by (13.16),

$$M(y^t F\mathfrak{S} y^{-t}) = (y^s G y^{-s})(y^{n+t} G^{-1} y^{-n-t}).$$

If $n + t \equiv s \bmod p$ this shows that $M \in \mathfrak{F}_2$. If $n + t \not\equiv s \bmod p$,

$$M(y^t F\mathfrak{S} y^{-t})(y^{n+t} F\mathfrak{S} y^{-n-t}) = (y^s G y^{-s})(y^{2n+t} G^{-1} y^{-2n-t}).$$

If $2n + t \equiv s \bmod p$ this shows that $M \in \mathfrak{F}_2$. Otherwise a third step is made and so on. But by (13.12), n and p are relatively prime. Hence there is a k such that $kn + t \equiv s \bmod p$, and so after k steps it results that $M \in \mathfrak{F}_2$. Hence H and LN determine the same element of μ_2. Now the argument applied to H can be applied to LN and so on until eventually it follows that H defines the same element of μ_2 as the identity. This shows that for the lens space (p, q), μ_2 is the identity.

CORNELL UNIVERSITY

Annals of Mathematics
Vol. 44, No. 4, October, 1943

SOME REMARKS ON SET THEORY

By P. Erdös

(Received February 1, 1943)

This paper contains a few disconnected results on the theory of sets.

I. Sierpinski[1] proved that under the assumption of the continuum hypothesis there exists a single valued function $f(x)$ whose inverse function is also single valued and which maps the sets of measure 0 into sets of first category and whose inverse function maps the sets of first category into sets of measure 0. He stated the problem[2] whether a function exists which has the above property and also the following one: It maps the sets of first category into sets of measure 0 and its inverse function maps the sets of measure 0 into sets of first category. Thus the function would interchange the sets of measure 0 and the sets of first category. We shall prove that such a function exists. Our proof will be very similar to that of Sierpinski: we will of course assume that the continuum hypothesis holds.

Construction of $f(x)$: It can be shown[3] that a transfinite sequence G_α of G_δ sets of measure 0 and a transfinite sequence F_α of F_σ sets of first category exists ($\alpha < \Omega_1$, Ω_1 is the first ordinal number of the third number class) having the following properties: 1) $G_1 \cup F_1 = R$, $G_1 \cap F_1 = \Lambda$ (R denotes the set of all real numbers), 2) every set of measure 0 is contained in some G_α and every set of first category is contained in some F_α 3) $A_\alpha = G_\alpha - G_\alpha \cap \bigcup_{\beta<\alpha} G_\beta$, $B_\alpha = F_\alpha - F_\alpha \cap \bigcup_{\beta<\alpha} F_\beta$ both have the power of the continuum, for every α. We evidently have $G_1 = \bigcup_{\alpha>1} B_\alpha$, $F_1 = \bigcup_{\alpha>1} A_\alpha$. Hence we can construct a function $f(x)$ in such a way that $f(A_\alpha) = B_\alpha$ for every $\alpha > 1$, and that $ff(x) = x$ for every x. The function $f(x)$ is clearly a single valued function whose inverse $f^{-1}(x)$ is also single valued. Since, in addition $f(x)$ coincides with its inverse, we have only to show that $f(x)$ maps both the sets of measure 0 onto sets of first category, and the sets of first category onto sets of measure 0. But both of these statements are obvious. For let G be any set of measure 0; by assumption $G \subset G_\alpha$ for some α and $f(G) \subset \bigcup_{\beta<\alpha} F_\beta$, which is a set of first category. Similarly let F be any set of first category; by assumption $F \subset F_\alpha$ for some α, and $f(F) \subset \bigcup_{\beta<\alpha} G_\beta$ which is a set of measure 0: This completes the proof.

II. Let m be a cardinal number. Two sets A and B in Euclidean space are called m-equivalent if they can be split into m summands $A = UA_\alpha$, $B = UB_\alpha$, $A_{\alpha_i} \cap A_{\alpha_j} = B_{\alpha_i} \cap B_{\alpha_j} = 0$, and $A_\alpha \cong B_\alpha$. (The sign $\cong$ denotes congruence.)

[1] W. Sierpinski, Fund. Math. Vol. 22, p. 276–278.

[2] Ibid.

[3] Ibid.

Banach and Tarski[4] proved that in three space any two sets containing open sets are finitely equivalent, and that on the line and the plane any two sets containing open sets are countably equivalent.

Professor Tarski[5] communicated to me the following result of Lindenbaum: There exist 2^c linear sets no two of which are countably equivalent. This result was never published, and Tarski does not remember the details of the proof. I have succeeded in proving that if m is any cardinal number $< c$, then there exist 2^c linear sets no two of which are m-equivalent. I do not know whether my proof differs from that of Lindenbaum, but I have thought it might be worth publishing, since the result has some interesting applications.

First we remark that it is easy to construct 2^n subsets of an infinite set A of power n such that the symmetric difference $(x - y) \cup (y - x)$ of any two subsets x and y has the power n. It is sufficient to divide A into n mutually exclusive subsets of power n, and to consider the unions of all these subsets.

Let now $\{a_\alpha\}$ be a Hamel base ($\alpha < \omega_\xi$, ω_ξ is the smallest ordinal belonging to the power of the continuum.) and let $A_\beta(\beta < \omega_\eta$, ω_η the smallest ordinal belonging to 2^c) be a family of subsets of this Hamel base such that the symmetric difference $(A_{\beta_1} - A_{\beta_2}) \cup (A_{\beta_2} - A_{\beta_1})$ has always the power c. Denote by U_β the set of real numbers of the form $\sum c_k a_k$ where the c_k are rational numbers and the a_k belong to A_β. Now we show that for $\beta_1 \neq \beta_2$ U_{β_1} and U_{β_2} are not m-equivalent. We can clearly assume that A_{β_2} contains c elements not contained in A_{β_1}. A being a set of numbers and x an arbitrary number, let us denote by $A + x$ the set of all numbers $z + x$ where z belongs to A. Also we denote by $A^{(y)}$ the reflection of A with respect to y. It suffices to show that if $\{x_{\mathfrak{S}}\}$ and $\{y_{\mathfrak{S}}\}$ are two sets of power m ($\mathfrak{S} < \omega_\delta$, ω_δ is the smallest ordinal number belonging to m) then the union of all the sets $U_{\beta_1} + x_{\mathfrak{S}}$, $U_{\beta_1}^{(y_{\mathfrak{S}})}$ does not contain U_{β_2}. And this is clear for if we denote by a_ξ the elements of the Hamel base necessary to express the $x_{\mathfrak{S}}$ and the $y_{\mathfrak{S}}$ (the power of the a_ξ is clearly $\leqq m$) our set $\bigcup U_{\beta_1} + x_{\mathfrak{S}}$, $U_{\beta_1}^{(y_{\mathfrak{S}})}$ can therefore be generated by the elements of A_{β_1} and by at most m other elements of the Hamel basis; while U_{β_2} is generated by the elements of A_{β_2}, and the latter set contains c elements which do not belong to A_{β_1}. This completes our proof.

III. A set B of real numbers is said to be of absolute measure 0 if it is finitely equivalent to a subset of an arbitrarily small interval. It is said to be of absolute measure α if for every ϵ it is finitely equivalent to a subset of an interval of length $\alpha + \epsilon$, and a subset of it is finitely equivalent with the interval of length $\alpha - \epsilon$.[6]

It is well known that the power of Lebesgue measurable sets mod null sets is of power c, but that the power of Lebesgue measurable sets is 2^c. Tarski[7]

---

[4] Banach and Tarski, Fund. Math. Vol. 6, p. 244–278.

[5] Oral communication.

[6] Tarski, Fund. Math. Vol. 30, p. 218–253. This paper contains the definition and all the properties used of absolute measure used in this proof.

[7] Oral communication.

posed the problem: What is the power of absolutely measurable sets mod sets of absolute measure 0? (It is of course clear the power of absolutely measurable sets is 2^c.)[8] We are going to prove that the power in question is 2^c.

First it is clear that it suffices to prove that the power of all sets in the interval (0, 1) mod sets of absolute measure 0 is of power 2^c. For if we take any set A in (0, 1) and translate it by 1, take its complement in (1, 2) denote it by B, then $A + B$ has absolute measure 1, hence if A_1 and A_2 are not congruent mod sets of absolute measure 0, $A_1 + B_1$ and $A_2 + B_2$ are also not congruent. This is a strong indication of the truth of our theorem, since it is well known that the power of all sets mod sets of Lebesgue measure 0 is also 2^c.

To prove our theorem it clearly suffices to show that there exist in the interval (0, 1) c disjoint sets whose absolute measure is not 0, for by taking all possible sums of these sets we clearly get 2^c sets no two of which are congruent mod sets of absolute measure 0.

Let now $\{a_\alpha\}$ be a Hamel base with $a_1 = 1$. Split it into c disjoint sets of power c. Denote these sets by V_β. We define the sets R_β as follows: $x \in R_\beta$ if and only if $0 \leqq x \leqq 1$ and $x = \sum_{i=1}^{k} c_i a_{\alpha_i}$, the c_i rational and different from 0 and $\alpha_1 < \alpha_2 < \cdots \alpha_k$, $a_{\alpha_k} V_\beta$. (For $i < k$ a_{α_i} does not have to belong to V_β.) We are going to prove that the disjoint sets R_β are not of absolute measure 0. In fact we shall show that R_β is not finitely equivalent with any subset of $(0, \frac{1}{2})$. For suppose that R_β is finitely equivalent with a subset of $(0, \frac{1}{2})$. This would mean that there exist sets $U_1, U_2, \cdots, U_r$ whose sum is the interval $(0, \frac{1}{2})$, and real numbers $x_1, x_2, \cdots x_j, y_1, y_2, \cdots y_l$, $k + l = r$, such that, R_β is contained in $\bigcup U_i + x_i$, $U_j^{(y_j)}$, the sets U_i are supposed to be mutually exclusive. Let a_k be the a_α of largest index which occurs in the representation of the x_i and y_j and denote by R'_β those elements of R'_β in whose representation the a_α of largest index has an index $> k$. Then if w is an element of R'_β and $w \in U_i + x_i$ there exists a $z \in U_i$ with $z = w - x_i$, hence $z \in R'_\beta$, also if $w \in U_j^{(y_j)}$ there again exists a $z \in U_j$ with $z = 2y_j - w$ hence $z \in R'_\beta$. Similarly if $z \in R'_\beta$ we have $z + x_i \in R'_\beta$ and $2y_j - z \in R'_\beta$. Thus we see that $(R'_\beta \cap U_i) + x_i = R'_\beta \cap (U_i + x_i)$ and $(R'_\beta \cap U_j)^{y_j} = R'_\beta \cap U_j^{y_j}$, hence we conclude that $R'_\beta \cap (0, \frac{1}{2})$ is finitely equivalent to $R'_\beta = R'_\beta \cap (0, \frac{1}{2}) \cup R'_\beta \cap (\frac{1}{2}, 1)$. On the other hand a translation by $\frac{a_1}{2} = \frac{1}{2}$ shows that $R'_\beta \cap (0, \frac{1}{2}) \cong (R'_\beta \cap (0, \frac{1}{2})) + \frac{1}{2} = R'_\beta \cap (\frac{1}{2}, 1)$. Thus $R'_\beta \cap (0, \frac{1}{2})$ would be finitely equivalent with $R'_\beta \cap (0, \frac{1}{2}) \cup (R'_\beta \cap (0, \frac{1}{2})) + \frac{1}{2}$. A general theorem of Lindenbaum and Tarski[9] shows that this is not possible, which completes our proof.

Sierpinski[10] constructed a set k of real numbers of power c whose complement has also power c, and such that if $k \cong k'$ then the power of $k' \cap (R - k)$ [as before R denotes the set of all real numbers] is $<c$. It is easy to see that if we define R_β as in III but remove the restriction $0 \leqq x \leqq 1$. Our set R_β has the required property.

---

[8] This statement follows from the fact that there exist sets of absolute measure 0 having power c.

[9] Lindenbaum and Tarski.

[10] W. Sierpinski, Fund. Math. Vol. 19, p. 22–28.

We are going to prove that Sierpinski's theorem can not be improved i.e. if m is a power $<c$ there exists a number x_0 such that $k + x_0 \cap R - k$ has power $\geqq m$. Let y_α be any set of real numbers of power m, and suppose that our theorem does not hold. Then both $k + y_\alpha \cap R - k$ and $(R - k) + y_\alpha \cap k$ have power $<m$ for all α. Therefore since $m^2 = m$ it is easy to see that there exists a $z \epsilon k$ and a $w \epsilon R - k$ such that $z + y_\alpha \epsilon k$ and $w + y_\alpha \epsilon R - k$ for all α. But then clearly $x_0 = w - z$ has the required property, which completes our proof.[11]

IV. Let $f(x)$ be a continuous function in the closed interval (0, 1). Denote by E the set for which

$$\overline{\lim_{h \to +0}} \frac{f(x + h) - f(x)}{h} < \infty.$$

Jarnik[12] proved that E is not countable. We are going to give a very simple proof that E is of power c. (It is easy to see that the complement of E is an F_σ, thus from the fact that E is not enumerable it immediately follows that E is of power c.)

Let x_0 be a number <1 for which $\overline{\lim} \frac{f(x + h) - f(x)}{h} < \infty$. We can of course assume that such a number exists. Let $N > \frac{f(1) - f(x_0)}{1 - x_0}$, and consider the set of numbers for which $\frac{f(y) - f(x_0)}{y - x_0} \geqq N$. Consider the greatest such y and denote it by y_N. Clearly $y_N < 1$. Hence evidently

$$\frac{f(y_N + h) - f(y_N)}{h} < N \quad \text{for} \quad h < 0.$$

Thus y_N belongs to E. Also we have $f(y_N) - f(x) = N(y_N - x)$, hence for $N_1 > N_2, y_{N_1} < y_{N_2}$. Thus the power of points y_N is c, which completes the proof.

Professor Anthony P. Morse communicated to me the following proof of Jarnik's theorem which he obtained some time ago: Choose k so that if we put $g(x) = f(x) - kx$ we shall have $g(0) > g(1)$. Now take any number c such that $g(0) > c > g(1)$. There clearly exists an x such that $g(x) = 0$. Let x_c be the largest such x. It is easily seen that

$$\overline{\lim_{h \to +0}} \frac{g(x_c + h) - g(x_c)}{h} \leqq 0$$

and hence

$$\overline{\lim_{h \to +0}} \frac{f(x_c + h) - f(x_c)}{a} \leqq k.$$

Thus x_c belongs to E. The power of points x_c is clearly equal to that of the continuum, which completes the proof.

UNIVERSITY OF PENNSYLVANIA

[11] This proof is due to Mr. P. Lax. Oral communication.

[12] Jarnik, Fund. Math. Vol. 23, p. 1–8.

ANNALS OF MATHEMATICS
Vol. 44, No. 4, October, 1943

CORRECTIONS TO TWO OF MY PAPERS

BY P. ERDÖS

(Received February 24, 1942)

In my paper "*On the divergence properties of the Lagrange interpolation polynomials,*" (Annals of Math. Vol. 42, (1941), p. 309–315) I stated that, if $x_0 = \cos \frac{p}{q} \pi$ (p and q odd), and the fundamental points of the interpolation are the roots of the Tchebicheff polynomial $T_n(x)$, then there exists a continuous function $f(x)$ such that $\lim L_n(f(x_0)) = \infty$.

Dr. Schönberg has pointed out that the proof there given is not correct. There is a trivial error in lemma 1; namely, it is possible that $x_i^{(m)} = x_j^{(n)}$. Nevertheless it is possible to save almost everything, practically without modifying the proof. We prove the following slightly weaker.

THEOREM. *There exists a continuous function $f(x)$ such that if $x_0 = \cos \frac{p}{q} \pi$, where p and q are odd, then* $\lim | L_n(f(x_0)) | = \infty$.

Proof. We need

LEMMA 1. *If $x_i^{(m)} \neq x_j^{(n)}$ then $| x_i^{(m)} - x_j^{(n)} | > \frac{1}{m^3}$ for $m \geqq n$.*

Proof. As in the paper.

Everything is now unchanged until the bottom of page 311. We have there

$$f(x) = \sum_{n=n_0}^{\infty} \epsilon_n \frac{f_n(x)}{\sqrt{\log n}}.$$

where $\epsilon_n = \pm 1$ and will be determined later; the definition of $f_n(x)$ is the same as in the paper.

$L_n(\varphi_2(x_0)) = 0$ still holds (p. 313 top). It suffices to show that, for $r > n$, $f_r(x_k^{(n)}) = 0$. And this is true, for otherwise either

$$x_l^{(r)} = x_k^{(n)},$$

which is impossible since $(2l - 1, r) = 1$, or we have

$$x_l^{(r)} \neq x_k^{(n)} \quad \text{and} \quad | x_l^{(r)} - x_k^{(n)} | < \frac{1}{2^{2^r}},$$

which does not hold by lemma 1.

Define now $\epsilon_n = \text{signum } L_n(\varphi_1(x_0))$; then clearly

$$| L_n(f(x_0)) | \leqq \left| L_n\left(\frac{\epsilon_n f_n(x_0)}{\sqrt{\log n}}\right) \right|$$

and the rest of the proof is unchanged.

At present I cannot decide whether a continuous function $f(x)$ exists such that $\lim L_n(f(x_0)) = \infty$, or whether a continuous $f(x)$ exists with $\lim L_n(f(x_0)) = a$, where $a \neq f(x_0)$.

Added in proof. By a more careful analysis, I can now show the following theorem: *Let E be any closed set, then there exists a continuous $f(x)$ such that the limit points of L_n $(f(x_0))$ is precisely the set E. The set E can consist of the point $+\infty$ alone.* This of course is a generalization of the result mentioned before.

In my paper "*On some asymptotic formulas in the theory of factorisatio numerorum*" (Annals of Math. Vol. 42, (1941) p. 989–993) the main theorem is stated incorrectly. The correct statement is as follows:

Let $1 < a_1 < a_2 < \cdots$ be a sequence of integers such that for some ρ, $\sum_{i=1}^{\infty} \frac{1}{a_i^{\rho}} = 1$ and $\sum \frac{\log a_i}{a_i^{\rho}}$ converges and not all the a_i's are powers of a_1. Denote by $f(n)$ the number of factorisations of n into the a_i's. We consider order in other words $a_1 a_2$ and $a_2 \cdot a_1$ are different factorisations. Also $f(1) = 1$. Denote $F(n) = \sum_{k=1}^{n} f(k)$. Then we have

$$F(n) = cn^{\rho}(1 + 0(1)).$$

The proof remains entirely unchanged: in fact this theorem is the one really proved in the paper.

It might be of some interest to investigate what happens if the conditions of our theorem are not satisfied. There are three cases: I. $\sum_{i=1}^{\infty} \frac{1}{a_i^k}$ diverges for all k. Then it is easy to see that $\lim \frac{F(n)}{n^k} = \infty$ for all k.

II. For all values of k for which $\sum_{i=1}^{\infty} \frac{1}{a_i^k}$ converges $\sum_{i=1}^{\infty} \frac{1}{a_i^k} < 1$. Clearly, there exists a ρ such that for every ϵ, $\sum_{i=1}^{\infty} \frac{1}{a_i^{\rho+\epsilon}}$ converges but $\sum \frac{1}{a_i^{\rho-\epsilon}}$ diverges. We can easily see that $\sum_{i=1}^{\infty} \frac{1}{a_i^{\rho}}$ converges and is <1. For if $\sum_{i=1}^{\infty} \frac{1}{a_i^{\rho}}$ diverged we would have, for sufficiently small ϵ, $\sum_{i=1}^{\infty} \frac{1}{a_i^{\rho+\epsilon}} > 1$; and since, for large k, $\sum_{i=1}^{\infty} \frac{1}{a_i^k} < 1$, there would exist a k_0 such that $\sum_{i=1}^{\infty} \frac{1}{a_i^{k_0}} = 1$—which contradicts the hypothesis.

Now we show that

$$\lim \frac{F(n)}{n^{\rho}} = 0 \tag{1}$$

and

$$\lim \frac{F(n)}{n^{\rho-\epsilon}} = \infty. \tag{2}$$

Suppose (1) does not hold. Write $\sum \frac{1}{a_i^{\rho}} = A < 1$ and $c = \limsup \frac{F(u)}{u^{\rho}}$.

We have

$$F(u) = \sum_{i=1}^{\infty} F\left(\frac{u}{a_i}\right) + 1,$$

so that

$$\frac{F(u)}{u^\rho} \leqq c \sum_{i=1}^{\infty} \frac{1}{a_i^\rho} + o(1) < Ac + o(1)$$

for sufficiently large u. This is possible only if $c = 0$. (2) can be shown by similar arguments.

III. There exists a ρ such that $\sum_{i=1}^{\infty} \frac{1}{a_i^\rho} = 1$, but $\sum_{i=1}^{\infty} \frac{\log a_i}{a_i^\rho}$ diverges. It seems likely that in this case

$$\lim \frac{F(n)}{n^\rho} = 0.$$

But I am only able to prove that

$$\varliminf \frac{F(n)}{n^\rho} = 0. \tag{3}$$

Suppose that (3) is not satisfied. Let the greatest lower bound of $\frac{F(n)}{(n+1)^\rho}$ be c $(c > 0)$. Choose k so large that

$$\sum_{i=1}^{k} \frac{\log a_i}{a_i^\rho} > \frac{2}{\rho c}.$$

Denote by $g(n)$ the number of the factorisations of n as the product of the a_i for $i \leqq k$, and let $G(n) = \sum_{u=1}^{n} g(u)$. Clearly for $m \leqq a_k$, $G(m) = F(m)$. Thus for $m \leqq a_k$

$$\frac{G(m)}{(m+1)^\rho} \geqq c.$$

Next we prove that for all m

$$\frac{G(n)}{(n+1)^{\rho'}} \geqq c \tag{4}$$

where $\sum_{i=1}^{k} \frac{1}{a_i^{\rho'}} = 1$, $(\rho' \leqq \rho)$.

Clearly (4) holds for all $n \leqq a_k$. We prove (4) by induction. Assume it for n: we shall prove it for $n + 1$. We have

$$G(n+1) = \sum_{i=1}^{k} G\left[\frac{n+1}{a_i}\right] + 1 \geqq c \sum_{i=1}^{k} \left(\left[\frac{n+1}{a_i}\right] + 1\right)^{\rho'} + 1.$$

Therefore

$$\frac{G(n+1)}{(n+2)^{\rho'}} \geqq c \sum_{i=1}^{k} \frac{1}{a_i^{\rho'}} = c,$$

which proves (4). Thus $\lim \frac{G(n)}{n^{\rho'}} \geqq c$. (We know from my paper that the limit exists.)

Put $h(s) = \sum_{i=1}^{k} \frac{1}{a_i^s}$; then clearly

$$\frac{1}{2 - h(s)} = \sum_{n=1}^{\infty} \frac{g(n)}{n^s}.$$

Therefore a simple calculation shows that if $\lim \frac{G(n)}{n^{\rho'}}$ exists the limit equals

$$\left(\rho' \sum_{i=1}^{k} \frac{\log a_i}{a_i^{\rho'}}\right)^{-1} < \frac{\rho c}{2\rho'} \leqq \frac{c}{2},$$

which proves (3). It is easy to construct sequences a_i, with $\sum_{i=1}^{\infty} \frac{1}{a_i^{\rho}}, = 1$, $\sum_{i=1}^{\infty} \frac{\log a_i}{a_i^{\rho}} = \infty$ and

$$\lim \frac{F(n)}{n^{\rho}} = 0. \tag{5}$$

But I can not prove that (5) holds for all such sequences a_i.

Professor Hille has given the following result. (Acta Arithmetica Vol. 2, p. 140): Let $p_1 < p_2 < \cdots$ be a sequence of primes, and let $a_1 < a_2 < \cdots$ be the integers composed of the p's. Denote by $f(n)$ the number of factorisations of n into the product of the a's, and let $F(n) = \sum_{m=1}^{n} f(m)$. If $\sum_{i=1}^{\infty} \frac{1}{a_i^{\rho}} = 1$ then $\lim \frac{F(n)}{n^{\rho}} = \left(\rho \sum_{i=1}^{\infty} \frac{\log a_i}{a_i}\right)^{-1}$. His proof (which uses the theorem of Wiener Ikehara) seems to apply only if $\sum_{i=1}^{\infty} \frac{\log a_i}{a_i^{\rho}} < \infty$. If (5) is always true in case iii, Hille's result would follow even if $\sum_{i=1}^{\infty} \frac{\log a_i}{a_i^{\rho}} = \infty$.

Recently I found in the literature a few results, which I proved in my paper "Elementary proof of some asymptotic formulas in the theory of partitions" (Annals of Math. Vol. 43). On p. 447 I prove the following result: Denote by $P_r(n)$ the number of partitions of n into powers of r then $\lim \frac{\log P_r(n)}{(\log n)^2} = \frac{1}{2 \log r}$. This result was proved by Mahler (London Math. Soc. Journal, Vol. 15, p. 123.) Mahlers proof is completely different from mine. He also obtains

$$c_1 \frac{r^{-\frac{1}{2}n(n-1)}(zr)^n}{n!} < P_r(rz) < c_2 \frac{r^{-\frac{1}{2}n(n-1)}(rz)^n}{n!}, \quad \text{where} \quad r^{n-1}n \leqq z < r^n(n+1).$$

On p. 448 I prove the following two results:

I. Let $a_1 < a_2 < \cdots$ be a sequence of integers of positive density α, the a's have common factor 1. Denote by $p(n)$ the number of partition of n into the a's. Then $\log p(n) \sim \pi\sqrt{\frac{2}{3}\alpha n}$.

II. Let $a_1 < a_2 < \cdots$ be a sequence of integers such that every large integer is the sum of different a's. Denote by $P(n)$ the number of partitions of n into different a's. Then $\log P(n) \sim \pi\sqrt{\frac{1}{3}\alpha n}$. Similar results were proved by K. Knopp (Schriften der Königsberger Gel. Ges. Math. und Nat. Klasse, 2 Jahr. Heft. 3 1925). His proofs are quite different from mine and are more complicated.

UNIVERSITY OF PENNSYLVANIA

ADDENDUM

BY EINAR HILLE

The objection raised by Dr. Erdös to formula (4.3) of my paper "A Problem in 'Factorisatio Numerorum' " is well founded. However, the results on pp. 139–140 are entirely correct if the basis P contains only a finite number of primes p_{i_ν}. When the basis is infinite it is necessary to assume that $\lim_{s\to\sigma_0} \zeta(s; P) > 2$ where $\zeta(s; P) = \prod_{\nu=1}^{\infty} [1 - p_{i_\nu}^{-s}]^{-1}$ and $\sigma_0 = \sigma_0(P)$ is the abscissa of convergence of the infinite product. This assumption implies that the equation $\zeta(s; P) = 2$ has a root $\rho(P)$ which exceeds σ_0. If this assumption is satisfied, formulas (3.8), (3.9), (4.1), (4.3), and (5.1) remain valid. If, instead, $\zeta(\sigma_0; P) = 2$ so that $\rho(P) = \sigma_0$, the Ikehara-Wiener theorem does not apply; the analysis breaks down completely and cannot be saved by assuming that $\zeta'(\sigma_0; P)$ is finite. Though formula (4.3) still makes sense, it is at best unproved. If $\zeta(\sigma_0; P) < 2$, the formula becomes meaningless and it is not enough to replace $\rho(P)$ by σ_0 since Erdös has proved [formula (1) above] that in this case $F(n) = o(n^{\sigma_0})$ while it is not necessarily true that $\zeta'(\sigma_0 ; P)$ is infinite.

YALE UNIVERSITY

Annals of Mathematics
Vol. 44, No. 4, October, 1943

ANALYTIC AND MEROMORPHIC CONTINUATION BY MEANS OF GREEN'S FORMULA

S. Bochner

(Received April 28, 1943)

Poincaré was the first to apply methods of potential theory to the study of analytic functions in several complex variables.[1] Using his method of "balayage" he proved for the entire space that meromorphic functional elements whose difference is holomorphic where they overlap can be merged into one meromorphic function. Later Cousin extended the result to the more general case of a polycylindrical domain, but he used the method of iterating the classical Cauchy integral in one complex variable.

In the present paper we will use a greatly simplified approach to Poincaré's original method, and our main purpose is to show that his theorem is closely connected with the theorem of Hartogs that every analytic function in several complex variables can be continued from the connected boundary of a domain into its whole interior. The connecting link is Green's formula for general harmonic functions and a related version of Cauchy's formula. The first application of this method to Hartogs' theorem is due to Fueter. However Fueter has a very involved approach to the method from the standpoint of quaternions and other Cayley number systems. Our starting point is the ordinary Green's formula, and the specialized results will follow from the fact that an arbitrary differential operator with constant coefficients, being commutative with the Laplacean, is also commutative with Green's integral which is the operational inverse to the Laplacean. This approach will yield old and new results in very general versions and will also permit us to extend Poincaré's theorem to functions on the torus.

Chapter I

Stokes' Theorem

1. We will consider a Euclidean space E_n, $n \geqq 3$. If A is a pointset then $\bar{A}$ denotes its closure and $\tilde{A}$ some neighborhood of $\bar{A}$ (no matter how small).

[1] For references to the theorem of Poincaré-Cousin see the article by H. Behnke and P. Tullen, Theorie der Funktionen mehrerer komplexen Veränderlichen, in Ergebnisse der Mathematik, volume III, pp. 64–67; for references to the theorem of Hartogs see the same article, p. 50, and two more recent papers by R. Fueter in Commentarii Mathematici Helvetici, volume 12 (1939), pp. 75–80, and volume 14 (1942), pp. 394–400.

Added in proof, Sept. 1943. Formula (53) of the present paper and a proof of theorem 5 based on it have just been published by Enzo Martinelli, Sopra una dimonstrazione di R. Fueter per una teorema di Hartogs, Comm. Helvetici, 15 (1942–3), submitted Jan. 1943. The present author may be permitted to state that these results have been presented by him in a Princeton graduate course in Winter 1940/41 and were subsequently incorporated, in a Princeton doctorate thesis (June 1941) by Donald C. May, entitled: An integral formula for analytic functions of k variables with some applications.

Virtually all our pointsets will be simplices or finite unions of disjoint simplices. Whenever a symbol like B, C, D will be introduced to denote one or several simplices then the same symbol will also be used to denote the corresponding pointset. If B (or C, D) is an m-dimensional simplex in E_n, then it will be always implicitly understood that B is a topological map of the closure $\bar{S}$ of the rectilinear fundamental simplex S in E_m and the mapping transformation is defined and of class one (with regard to differentiability) in $\bar{S}$. All functions introduced will be likewise of class one unless more restrictive assumptions are explicitly stated. It will be sufficient for our purposes to express Stokes' theorem as the formula

$$\int_B d\varphi\, d\psi_1 \cdots d\psi_{m-1} = \int_C \varphi\, d\psi_1 \cdots d\psi_{m-1}. \tag{1}$$

The scalars $\varphi, \psi_1, \cdots, \psi_{m-1}$ are defined in $\bar{B}$ and C is the set of the $(m-1)$-dimensional faces of B. The simplex B is oriented and the orientation of C is related to the orientation of B by the well known combinatorial rule. By algebraic addition the formula extends readily from a simplex to a (finite) simplicial manifold. In particular, if D is a simplicial domain in E_n with boundary $\bar{B}$, and X_α, $\alpha = 1, \cdots, n$, is a vector field in $\bar{D}$, then

$$\int_D \sum_{\alpha=1}^{n} \frac{\partial X_\alpha}{\partial \xi_\alpha}\, dv_\xi = \int_B \sum_{\alpha=1}^{n} X_\alpha \sigma^\alpha, \tag{2}$$

where

$$dv_\xi = d\xi_1 \cdots d\xi_n \tag{3}$$

$$\sigma^\alpha = (-1)^{\alpha-1} d\xi_1 \cdots d\xi_{\alpha-1} d\xi_{\alpha+1} \cdots d\xi_n\,; \tag{4}$$

or, omitting the symbol of summation with respect to α,

$$\int_D \frac{\partial X_\alpha}{\partial \xi_\alpha}\, dv_\xi = \int_B X_\alpha \sigma^\alpha. \tag{5}$$

The symbols (3), (4) are external differential forms, and the symbol (3) will be the ordinary Euclidean volume element if D is cooriented with E_n; this will be always assumed.

2. We consider m-dimensional simplices $B_1, \cdots, B_s$ and $(m-1)$-dimensional simplices $C_1, \cdots, C_r$, each in a fixed orientation and all disjoint, and we assume that each C_p is a face of one or several B_q and that each $(m-1)$-dimensional face of each B_q occurs among the C_p. We introduce the incidence number ϵ_{pq} which is 0 if C_p is not a face of B_q and otherwise is $+1$ or -1 depending on whether the given orientation of C_p coincides with that of B_q or not.

If we put

$$B = B_1 + \cdots + B_s, \qquad C = C_1 + \cdots + C_r, \tag{6}$$

if the functions $\psi_1, \cdots, \psi_{m-1}$ are scalars in $\tilde{B}$, if $\varphi_q(\xi)$ is given in $\tilde{B}_q$, $q = 1, \cdots, s$, and if we put

$$\varphi^p(\xi) = \sum_{q=1}^{s} \epsilon_{pq} \varphi_q(\xi), \qquad p = 1, \cdots r, \tag{7}$$

then (1) implies

$$\sum_{p=1}^{r} \int_{C_p} \varphi^p \, d\psi_1 \cdots d\psi_{m-1} = \sum_{q=1}^{s} \int_{B_q} d\varphi_q \, d\psi_1 \cdots d\psi_{m-1} . \tag{8}$$

In order to assimilate relation (8) to the original relation (1) we introduce the concept of a *conglomerate function.* We take any functional elements φ_q, $q = 1, \cdots, s$, each defined separately on B_q (or $\tilde{B}_q$) without any interrelation, and we call the whole set a conglomerate function on B (or $\tilde{B}$); we denote it either by φ or more explicitly by $\{\varphi_q\}$. If $\{\varphi_q\}$ is defined on $\tilde{B}$, then the functional elements (7) give rise to a new conglomerate function, the latter defined on $\tilde{C}$. The new function will be called the *saltus* of φ on C. We can now interpret (1) to mean (8) if φ in $d\varphi d\psi_1 \cdots d\psi_{m-1}$ is any conglomerate function and the other φ in $\varphi d\psi_1 \cdots d\psi_{m-1}$ is its saltus. The concept of conglomerate functions unifies the common notions of "discontinuity" and "boundary value." In fact, if B is a simplicial manifold, then on any "internal" C_q which separates simplices B_a, B_b, the saltus of φ is $\pm(\varphi_a - \varphi_b)$ and on any boundary face C_q which borders on only one B_c, the saltus is $\pm\varphi$ itself.

3. We are now going to state a lemma of a special nature. If B and C are defined as before, and φ is a conglomerate function on $\tilde{B}$ then for $m \leqq n - 1$ we will investigate the change of the integral

$$\int_B \varphi \, d\psi_1 \cdots d\psi_m$$

for a small deformation of B. It will suffice to consider a deformation in which only one vertex (and the faces containing it) are involved. We denote this vertex by A_0 and its deformed image by A_0'. Consider all those among the simplices $B_1, B_2, \cdots$ which contain the vertex A_0 and denote them by $B_1, \cdots, B\mu$. Similarly denote by $C_1, \cdots, C_\lambda$ all faces containing A_0. By B_q', $q = 1, \cdots, \mu$, we denote the deformed image of B_q, by B_p'' we denote the m-dimensional complex which is the combinatorial product of A_0' and C_p, in this order of combination, and finally we denote by S_q the $(m + 1)$-dimensional simplex which is the product of A_0' and B_q. We now claim that the difference

$$\sum_{q=0}^{\mu} \int_{B_q} \varphi_q \, d\psi_1 \cdots d\psi_m - \sum_{q=0}^{\mu} \int_{B_q'} \varphi_q \, d\psi_1 \cdots d\psi_m \tag{9}$$

has the value

$$\sum_{q=0}^{\mu} \int_{S_q} d\varphi_q \, d\psi_1 \cdots d\psi_m + \sum_{p=0}^{\lambda} \int_{B_p''} \varphi^p \, d\psi_1 \cdots d\psi_m , \tag{10}$$

where $\{\varphi^p\}$ on $\tilde{C}$ is the saltus of $\{\varphi_q\}$ on $\tilde{B}$.

The proof consists in a simple combinatorial analysis. Consider any fixed simplex B_1 with vertex A_0 and write it combinatorially as $A_0A_1A_2 \cdots A_m$ where $A_1, \cdots, A_m$ are other vertices. We now introduce the $(m+1)$-dimensional simplex S_1, namely the simplex

$$A_0'A_0A_1 \cdots A_m. \tag{11}$$

Its m-dimensional faces in natural orientation are the sum of

$$A_0A_1A_2 \cdots A_m - A_0'A_1A_2 \cdots A_m \tag{12}$$

and

$$A_0'A_0A_2 \cdots A_m - A_0'A_0A_1A_3 \cdots A_m + \cdots \pm A_0'A_0A_1 \cdots A_{m-1}. \tag{13}$$

Now, (12) is simply $B_1 - B_1'$. As for (13), its first term is the product of A_0' with $A_0A_2 \cdots A_m$. The latter is the negative of $\epsilon_{1p}C_p$, for some p, where ϵ_{1p} is our previous incidence coefficient. Thus the first term in (13) is $-\epsilon_{1p}B_p''$ for some p. The same holds for any other term in (13). It is always $-\epsilon_{1p}B_p''$ for some p, the sign being always negative whether it was originally positive or negative in (13). The equality of (9) and (10) follows now by applying Stokes' formula to S_1 and its boundary, and to any other S_q and its boundary, and then adding up over those S_q which correspond to all those B_q which contain the given vertex A_0.

Chapter II

Analytic Functions of Real Variables

4. Green's formula. The most familiar general version of Green's formula is

$$\begin{matrix}\gamma_n f(x) \\ 0\end{matrix} = \int_B \left(f\frac{\partial G}{\partial \xi_\alpha} - G\frac{\partial f}{\partial \xi_\alpha}\right)\sigma^\alpha - \int_D (f\cdot\Delta G - G\cdot\Delta f)\, dv_\xi, \tag{14}$$

and we are adding the kindred formula

$$\begin{matrix}\gamma_n f(x) \\ 0\end{matrix} = \int_B f\frac{\partial G}{\partial \xi_\alpha}\sigma^\alpha - \int_D \left(f\cdot\Delta G + \frac{\partial f}{\partial \xi_\alpha}\frac{\partial G}{\partial \xi_\alpha}\right) dv_\xi. \tag{15}$$

The function G is

$$\frac{1}{2-n}\left(\sum_\alpha(\xi_\alpha - x_\alpha)^2\right)^{\frac{2-n}{2}} + H(\xi, x), \tag{16}$$

where $H(\xi, x)$ is analytic in $E_n \times E_n$; the function f is a function on $\bar{D}$, where D is a simplicial domain with boundary $\bar{B}$; the symbol Δ denotes the Laplacean

$$\Delta = \frac{\partial^2}{\partial \xi_1^2} + \cdots + \frac{\partial^2}{\partial \xi_n^2}; \tag{17}$$

the lower value 0 occurs when x is a point of $E_n - \bar{D}$ whereas the upper value $\gamma_n f(x)$ occurs when x is a point of D; and γ_n is the constant

$$\frac{2\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)}, \tag{18}$$

which constant is the $(n-1)$-dimensional volume of the sphere $\xi_1^2 + \cdots + \xi_n^2 = 1$. For x outside $\bar{D}$ our formulas are direct consequences of (5) if $X_\alpha = f\frac{\partial G}{\partial \xi_\alpha} - G\frac{\partial f}{\partial \xi_\alpha}$, or $X_\alpha = f\frac{\partial G}{\partial \xi_\alpha}$, and for x in D they are obtained from the same values X_α by first excising a small sphere with center at x and then letting the radius of the sphere tend to 0.

If $H(\xi, x) = 0$ in (16), that is

$$G(\xi, x) \equiv \frac{1}{2-n}\Big(\sum_\alpha(\xi_\alpha - x_\alpha)^2\Big)^{\frac{2-n}{2}} \tag{19}$$

then G is harmonic, that is a solution of $\Delta G = 0$. If in addition $f(\xi)$ is also harmonic then (14) reduces to

$$\int_B \left(f\frac{\partial G}{\partial \xi_\alpha} - G\frac{\partial f}{\partial \xi_\alpha}\right)\sigma^\alpha = \gamma_n f(x) \quad \text{or } 0. \tag{20}$$

Until further notice, the symbol G will always denote the Newtonian kernel (19).

5. Green's integral. We now consider a fixed set of simplices (6) and we specify that the dimension of B shall be $n-1$. We will call a conglomerate function $\{f_q\}$ on $\tilde{B}$ harmonic if each f_q is harmonic on $\tilde{B}_q$. With such a conglomerate harmonic function $f(\xi)$ we now set up the integral

$$F(x) = \frac{1}{\gamma_n}\int_B \left(f\frac{\partial G}{\partial \xi_\alpha} - G\frac{\partial f}{\partial \xi_\alpha}\right)\sigma^\alpha. \tag{21}$$

The complement $E_n - \bar{B}$ decomposes into a finite number of domains $D_1, \cdots D_t$, D_∞. The domain D_∞ is unbounded, the others are bounded. The bounded ones may be absent; this will happen if $\bar{B}$ does not decompose the space E_n. The union $D_1 + \cdots + D_t + D_\infty$ will be denoted by D, and we will treat D as if it were a union of simplices. Integral (21) defines a harmonic function $F_\tau(x)$ in each D_τ, $\tau = 1, \cdots, t, \infty$. In this sense, $\{F_\tau(x)\}$ is a harmonic conglomerate function on D.

THEOREM 1. *If $f(\xi)$ is a harmonic conglomerate function on $\tilde{B}$, then each $F_\tau(x)$ can be continued analytically across each B_q bordering on D_τ, and the saltus of $F(x)$ on B_q is $f_q(x)$.*

For the proof it may be assumed that B is just one simplex. We deform B into another simplex B' with the same boundary as B in such a way that $B' - B$ shall be part of the boundary of a simplicial domain D. The "side" of B on

which D is situated will be termed "positive," the other "negative." If x lies on the negative side of B it lies outside $\bar{D}$ and thus by (20) we have

$$F(x) = \frac{1}{\gamma_n}\int_{B'}\left(f\frac{\partial G}{\partial \xi_\alpha} - G\frac{\partial f}{\partial \xi_\alpha}\right)\sigma^\alpha. \tag{22}$$

Integral (22) however can be continued across B into D. Denoting this continuation by $F^+(x)$ we obviously have for x in D the relation

$$F^+(x) - F(x) = \frac{1}{\gamma_n}\int_{B'-B}\left(f\frac{\partial G}{\partial \xi_\alpha} - G\frac{\partial f}{\partial \xi_\alpha}\right)\sigma^\alpha. \tag{23}$$

However, again by (20), this is $f(x)$, and thus the theorem is proven.

THEOREM 2. *If also the saltus of $f(\xi)$ vanishes identically on $\tilde{C}$, then each F_τ can be continued into $\tilde{D}$.*

More generally, if a small deformation of a vertex A_0 of B enlarges a component D_τ into a domain D'_τ, and if the saltus of $f(\xi)$ vanishes on all those $\tilde{C}_p$ which contain the vertex A_0, then $F_\tau(x)$ can be continued from D_τ into D'_τ.

If we introduce the symbols B_q, B'_q, B''_p and S_q as in section 3, then by section 3 the expression

$$\sum_{q=0}^{\mu}\int_{B_q}\left(f_q\frac{\partial G}{\partial \xi_\alpha} - G\frac{\partial f_q}{\partial \xi_\alpha}\right)\sigma^\alpha - \sum_{q=0}^{\mu}\int_{B'_q}\left(f_q\frac{\partial G}{\partial \xi_\alpha} - G\frac{\partial f_q}{\partial \xi_\alpha}\right)\sigma^\alpha \tag{24}$$

is the sum of

$$\sum_{q=0}^{\mu}\int_{S_q}(f_q\cdot\Delta G - G\cdot\Delta f_q)\,dv_\xi \tag{25}$$

and

$$\sum_{p=0}^{\lambda}\int_{B''_p}\left(f^p\frac{\partial G}{\partial \xi_\alpha} - G\frac{\partial f^p}{\partial \xi_\alpha}\right)\sigma^\alpha. \tag{26}$$

Now, (24) is $\gamma_n(F(x) - F'(x))$ where the integral $F'(x)$ is what will become of $F(x)$ if B is replaced by its deformation B'. The sum (25) is naturally 0, and (26) is 0 if $f^p = 0$ on B''_p, that is if the saltus of f vanishes on $\tilde{C}_p$. Thus $F(x) = F'(x)$, but $F'(x)$ exists in D'_τ, and this proves our assertion.

6. Analytic continuation. We are now viewing the expression (21) as an operation

$$F(x) = Lf(\xi) \tag{27}$$

which transform a function in $\tilde{B}$ into a harmonic function in D. We claim that this operation is commutative with partial derivation.

THEOREM 3. *If $f(\xi)$ is harmonic in $\tilde{B}$ then*

$$L\frac{\partial f}{\partial \xi_\beta} = \frac{\partial}{\partial x_\beta}F(x), \tag{28}$$

provided the saltus of $f(\xi)$ and of $\frac{\partial f}{\partial \xi_1}, \cdots, \frac{\partial f}{\partial \xi_n}$ each vanishes on C.

More generally if the saltus of $f(\xi)$ and its partial derivatives of order $\leq N$ vanishes on C, then

$$L\Lambda f = \Lambda L f \tag{29}$$

where

$$\Lambda f \equiv \sum_{\nu_1+\cdots+\nu_n} a_{\nu_1} \cdots {}_{\nu_n} \frac{\partial^{\nu_1+\cdots+\nu_n} f}{\partial \xi_1^{\nu_1} \cdots \partial \xi_n^{\nu_n}} \tag{30}$$

with constant coefficients a.

REMARK. It should be noted that in Theorem 2, the saltus of f has to vanish identically on the open set $\check{C}$, whereas in Theorem 3 the saltuses have only to satisfy the *boundary condition of vanishing on the* $(n - 2)$*-dimensional set* C.

PROOF. Obviously (29) follows from (28) by induction on N. In addition to the symbols σ^α as defined by (4), we now introduce symbols $\sigma^{\alpha\beta}$ in the following way. For $\alpha < \beta$ we put

$$\sigma^{\alpha\beta} = (-1)^{\alpha+\beta-1} d\xi_1 \cdots d\xi_{\alpha-1} d\xi_{\alpha+1} \cdots d\xi_{\beta-1} d\xi_{\beta+1} \cdots d\xi_n;$$

for $\alpha = \beta$ we put $\sigma^{\alpha\beta} = 0$; and for $\alpha > \beta$ we put $\sigma^{\alpha\beta} = -\sigma^{\beta\alpha}$. The basis for our proof is the identity

$$\sum_{\alpha=1}^{n} \left(f \frac{\partial^2 G}{\partial \xi_\beta \partial \xi_\alpha} + \frac{\partial f}{\partial \xi_\beta} \frac{\partial G}{\partial \xi_\alpha} \right) \sigma^\alpha = \left(f \cdot \Delta G + \sum_{\alpha=1}^{n} \frac{\partial f}{\partial \xi_\alpha} \frac{\partial G}{\partial \xi_\alpha} \right) \sigma^\beta - \sum_{\alpha=1}^{n} d\left(f \frac{\partial G}{\partial \xi_\alpha} \right) \sigma^{\beta\alpha}. \tag{31}$$

From this we deduce that the quantity

$$\sum_{\alpha=1}^{n} \left(\frac{\partial f}{\partial \xi_\beta} \frac{\partial G}{\partial \xi_\alpha} - G \frac{\partial^2 f}{\partial \xi_\beta \partial \xi_\alpha} + f \frac{\partial^2 G}{\partial \xi_\beta \partial \xi_\alpha} - \frac{\partial G}{\partial \xi_\beta} \frac{\partial f}{\partial \xi_\alpha} \right) \sigma^\alpha \tag{32}$$

equals the quantity

$$(f \cdot \Delta G - G \cdot \Delta f) \sigma^\beta - \sum_{\alpha=1}^{n} d\varphi_\alpha \cdot \sigma^{\beta\alpha}, \tag{33}$$

where

$$\varphi_\alpha = f \frac{\partial G}{\partial \xi_\alpha} - G \frac{\partial f}{\partial \xi_\alpha}. \tag{34}$$

Now, since $\dfrac{\partial G}{\partial x_\beta} = -\dfrac{\partial G}{\partial \xi_\beta}$, the difference

$$\gamma_n \left(L \frac{\partial f}{\partial \xi_\beta} - \frac{\partial}{\partial x_\beta} F \right)$$

is the integral of (32) over B. By the equality of (32) and (33) this is

$$- \int_B \sum_{\alpha=1}^{n} d\varphi_\alpha \cdot \sigma^{\beta\alpha}$$

and by section 2 this will vanish if the saltus of each φ_α vanishes on C. By (34), it is enough to assume that $f, \dfrac{\partial f}{\partial \xi_1}, \cdots, \dfrac{\partial f}{\partial \xi_n}$ has vanishing saltus each.

We note a natural consequence of Theorem 3.

COROLLARY. *If the conglomerate function* $f(\xi)$ *is harmonic and the solution of* $\Lambda f = 0$ *and if the saltuses of* f *and of its partial derivatives of order* $\leqq N$ *vanish on* C *then* $F(x)$ *is likewise harmonic and a solution of* $\Lambda F = 0$.

We will say that a differential operator ΛF of type (30) has the *uniqueness property* if any analytic solution of $\Lambda F = 0$ which is defined in the complement of a bounded domain and vanishes at infinity is identically 0. Such an operator is for instance

$$\frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_m^2} \tag{35}$$

for $m < n$. In fact, if a function $F(x_1, \cdots, x_m, x_{m+1}, \cdots, x_n)$ is defined in the complement of a bounded domain in E_n then there exists an $(n - m)$-dimensional neighborhood of values $(x_{m+1}^0, \cdots, x_n^0)$, such that for each of those values, $F(x_1, \cdots, x_m, x_{m+1}, \cdots, x_n)$ is defined in

$$-\infty < x_\mu < \infty, \qquad \mu = 1, \cdots, m.$$

By our assumption it is a harmonic function in the entire $(x_1, \cdots, x_m)$-space and vanishes at infinity. Hence it is identically 0 in the latter variables; but being analytic in all n variables, it vanishes identically in all variables.

THEOREM 4. *If a conglomerate function* $f(\xi)$ *is harmonic in* $\tilde{B}$ *and the solution of an equation* $\Lambda f = 0$ *with uniqueness property; and if the saltuses of* $f(\xi)$ *and of its partial derivatives of order* $\leqq N$ *vanish on* C *then the following propositions hold.*

(i) *If the component* D_∞ *of* $E_n - \bar{B}$ *borders on a simplex* B_p *from both sides then* $f_p(x) = 0$.

(ii) *In particular, if* B *does not decompose the space then* $f(x) = 0$.

(iii) *If a simplex* B_p *separates* D_∞ *from* D_τ *then* $f_p(x)$ *can be continued into all of* D_τ.

(iv) *In particular, if* $\bar{B}$ *is the connected boundary of a bounded domain* D *then* $f(x)$ *can be continued (uniquely) into all of* D.

PROOF. *Ad* (i) *and* (ii). By the hypotheses of the theorem, $F_\infty(x)$ vanishes. But, by Theorem 3, in crossing over from D_∞ into D_∞ via B_p, $F(x)$ jumps by $\pm f_p(x)$. Hence $f_p(x) = 0$.

Ad (iii) *and* (iv). On B_p, $F(x)$ has the saltus $f_p(x)$. Therefore $F_\tau(x) - F_\infty(x) = \pm f_p(x)$. However, $F_\infty(x) = 0$. Thus, near B_p, $F_\tau(x)$ coincides with $\pm f_p(x)$. However, $F_\tau(x)$ exists in all of $\tilde{D}_\tau$. Therefore $f_p(x)$ exists in all of D_τ.

Now consider the space E_{2k}, $k \geqq 2$, of the real variables $x_1, y_1, \cdots, x_k, y_k$ and put $z_\alpha = x_\alpha + iy_\alpha$. If $f(z_1, \cdots, z_k)$ is an analytic function of the complex variables $z_1, \cdots, z_k$, then in particular its real and imaginary parts each satisfy the system of equations

$$\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial y_1^2} = 0, \cdots, \qquad \frac{\partial^2}{\partial x_k^2} + \frac{\partial^2}{\partial y_k^2} = 0. \tag{36}$$

Each of these equations has the uniqueness property. Also a simultaneous solution of them is harmonic in $(x_1, y_1, \cdots, x_k, y_k)$. Hence we obtain the following special case of theorem 4, (iv).

THEOREM 5. *If an analytic function of several complex variables is defined in the connected boundary of a bounded domain, then it can be continued uniquely into all of the domain.*

CHAPTER III

Analytic Functions of Complex Variables

7. Complex space. We will now consider systematically the space E_{2k} of k complex variables $\zeta_\alpha = \xi_\alpha + i\eta_\alpha$. All occurring functions will be likewise complex-valued. In particular a harmonic function will be one whose real and imaginary part are each harmonic. Replacing the pair of real variables ξ, η by the conjugate complex quantities $\zeta = \xi + i\eta$, $\bar{\zeta} = \xi - i\eta$ we have

$$d\bar{\zeta}d\zeta = 2id\xi d\eta \tag{37}$$

and therefore

$$d\bar{\zeta}_1 d\zeta_1 \cdots d\bar{\zeta}_k d\zeta_k = (2i)^k d\xi_1 d\eta_1 \cdots d\xi_k d\eta_k . \tag{38}$$

If any two of the $2k$ real factors on the right side of (38) are interchanged, the symbol changes its algebraic sign. However, if any two blocks of terms $d\bar{\zeta}_\alpha d\zeta_\alpha$, $d\bar{\zeta}_\beta d\zeta_\beta$ are interchanged the sign is not altered. It will be worth while to introduce the differential form

$$\omega = d\bar{\zeta}_1 d\zeta_1 + \cdots + d\bar{\zeta}_k d\zeta_k . \tag{39}$$

It is an "algebraic" area element, and *not* the scalar square of a length element. As for its (external) powers $\omega^2 = \omega \cdot \omega$, $\omega^3 = \omega^2 \cdot \omega$, $\cdots$, it can be easily seen that

$$\omega^k = k! d\bar{\zeta}_1 d\zeta_1 \cdots d\bar{\zeta}_k d\zeta_k \tag{40}$$

and that, for each α,

$$d\bar{\zeta}_\alpha d\zeta_\alpha \omega^{k-1} = \frac{1}{k} \omega^k. \tag{41}$$

As for the theorem of Stokes, we note that (1) remains valid for all occurring functions being complex-valued. In particular, if D is a domain with boundary B, then

$$\frac{1}{k} \int_D \frac{\partial X_\alpha}{\partial \bar{\zeta}_\alpha} \omega^k = \int_B X_\alpha d\zeta_\alpha \omega^{k-1} \tag{42}$$

and

$$-\frac{1}{k} \int_D \frac{\partial Y_\alpha}{\partial \zeta_\alpha} \omega^k = \int_B Y_\alpha d\bar{\zeta}_\alpha \omega^{k-1} . \tag{43}$$

We now put

$$V = 2G \tag{44}$$

where G is the function (16). Thus in the special case (19) we have

$$V = \frac{1}{1-k}\left(\sum_{\alpha} (\zeta_\alpha - z_\alpha)(\bar\zeta_\alpha - \bar z_\alpha)\right)^{1-k} \tag{45}$$

$$\frac{\partial V}{\partial \zeta_\alpha} = \frac{\bar\zeta_\alpha - z_\alpha}{\left(\sum_\beta |\zeta_\beta - z_\beta|^2\right)^k}. \tag{46}$$

In the general case (16), on putting

$$X_\alpha = f\frac{\partial V}{\partial \zeta_\alpha}, \qquad Y_\alpha = V\frac{\partial f}{\partial \bar\zeta_\alpha}$$

we obtain, as analogues to (14) and (15) the formulas

$$\begin{aligned} \begin{matrix}(2\pi i)^k f(z, \bar z)\\ 0\end{matrix} &= \int_B \left(f\frac{\partial V}{\partial \zeta_\alpha}\,d\zeta_\alpha + V\frac{\partial f}{\partial \bar\zeta_\alpha}\,d\bar\zeta_\alpha\right)\omega^{k-1} \\ &\quad - \frac{1}{k}\int_D \left(f\frac{\partial^2 V}{\partial \zeta_\alpha \partial \bar\zeta_\alpha} - V\frac{\partial^2 f}{\partial \zeta_\alpha \partial \bar\zeta_\alpha}\right)\omega^k \end{aligned} \tag{47}$$

and

$$\begin{matrix}(2\pi i)^k f(z, \bar z)\\ 0\end{matrix} = \int_B f\frac{\partial V}{\partial \zeta_\alpha}\,d\zeta_\alpha\,\omega^{k-1} - \frac{1}{k}\int_D \left(f\frac{\partial^2 V}{\partial \zeta_\alpha \partial \bar\zeta_\alpha} + \frac{\partial f}{\partial \bar\zeta_\alpha}\frac{\partial V}{\partial \zeta_\alpha}\right)\omega^k. \tag{48}$$

Since

$$4\frac{\partial^2}{\partial\zeta\partial\bar\zeta} = \frac{\partial^2}{\partial\xi^2} + \frac{\partial^2}{\partial\eta^2}$$

we have

$$4\sum_{\alpha=1}^{k}\frac{\partial^2}{\partial\zeta_\alpha\partial\bar\zeta_\alpha} = \sum_{\alpha=1}^{k}\left(\frac{\partial^2}{\partial\xi_\alpha^2} + \frac{\partial^2}{\partial\eta_\alpha^2}\right) \equiv \Delta. \tag{49}$$

Thus, since V is harmonic, if we also assume that f is harmonic, then (46) leads to

$$\int_B \left(f\frac{\partial V}{\partial \zeta_\alpha}\,d\zeta_\alpha + V\frac{\partial f}{\partial \bar\zeta_\alpha}\,d\bar\zeta_\alpha\right)\omega^{k-1} = (2\pi i)^k f(z, \bar z) \quad or \quad 0, \tag{50}$$

which is the analogue to (20). It should be observed that (50) is genuinely different from (20). All that can be said is that for real f, relation (50) is "the real part" of relation (20). The difference will be significant for general point-sets B, and will in fact lead to an improvement of Theorem 4.

8. Cauchy's formula. Now, if in particular f is an analytic function in $z_1, \cdots, z_k$ alone, thus

$$\frac{\partial f}{\partial \bar{z}_\alpha} = 0, \qquad \alpha = 1, \cdots, k, \tag{51}$$

(which assumption in particular implies that f is harmonic) then (50) reduces to the very simple formula

$$f(z) \text{ or } 0 = \frac{1}{(2\pi i)^k} \int_B f \frac{\partial V}{\partial \zeta_\alpha} d\zeta_\alpha \omega^{k-1}. \tag{52}$$

If we write the formula in full we obtain

THEOREM 6. *If $f(z)$ is analytic in $\bar{D}$ then the formula*

$$f(z) = \frac{1}{(2\pi i)^k} \int_B \frac{f(\zeta)(\bar{\zeta}_\alpha - \bar{z}_\alpha)\, d\zeta_\alpha \omega^{k-1}}{(\sum_\beta (\zeta_\beta - z_\beta(\bar{\zeta}_\beta - \bar{z}_\beta))^k} \tag{53}$$

holds for z in D.

This is a generalization of Cauchy's formula in one variable, however with the peculiarity that in the cases $k \geqq 2$ we are considering, the integrand contains $\bar{z}_\beta$ explicitly. Thus to start with it is not at all evident that the integral will be independent of $\bar{z}_\beta$ let alone be $f(z)$.

Now take a point $(z_1, \cdots, z_k)$ in D and replace in the integral (53) the conjugate point $(\bar{z}_1, \cdots, \bar{z}_k)$ by a point $(z_1', \cdots, z_k')$ in the vicinity of the conjugate point. The right side of (53) will then be a function

$$F(z_1, \cdots, z_k; z_1', \cdots, z_k')$$

in $2k$ complex variables. By Theorem 6, the function

$$F(z_1, \cdots, z_k; z_1', \cdots, z_k') - f(z_1, \cdots, z_k) \tag{54}$$

vanishes on the manifold $z_1' = \bar{z}_1, \cdots, z_k' = \bar{z}_k$. This implies that (54) vanishes identically, that is

$$f(z_1, \cdots, z_k) \equiv F(z_1, \cdots, z_k; z_1', \cdots, z_k').$$

Now assume that D contains the origin, and let $z_1', \cdots, z_k'$ be the origin. Then, at least for points z near the origin we have

$$f(z) = \frac{1}{(2\pi i)^k} \int_B \frac{f(\zeta)\bar{\zeta}_\alpha\, d\zeta_\alpha \omega^{k-1}}{[(\zeta_1 - z_1)\bar{\zeta}_1 + \cdots + (\zeta_k - z_k)\bar{\zeta}_k]^k}. \tag{55}$$

This formula is much closer to Cauchy's one-variable formula than (53) is, however the integral (55) does not necessarily converge in all of D. The domain of convergence can be described geometrically, but the description would be of no consequence in the present context.

9. Cauchy's integral. As an analogue to (21) we now introduce the integral

$$F(z, \bar{z}) = \frac{1}{(2\pi i)^k} \int_B f(\zeta) \frac{\partial V}{\partial \zeta_\alpha} d\zeta_\alpha \omega^{k-1} \tag{56}$$

for an analytic conglomerate function $f(\zeta_1, \cdots, \zeta_k)$ in $\tilde{B}$. We first of all have the following analogue to Theorems 1 and 2.

THEOREM 7. *The saltus of the conglomerate function* (56) *on B is $f(z)$.*

Also if the saltus of $f(\zeta)$ vanishes identically on all $\tilde{C}_p$ containing a vertex A_0 of B, then a small deformation of A_0 leads to analytic continuation of $F_\tau(z, \bar{z})$ from D_τ into D'_τ.

The following theorem is an analogue to Theorem 3 with

$$\Lambda f \equiv \frac{\partial f}{\partial \bar{z}_\beta} = \frac{\partial f}{\partial x_\beta} + i\,\frac{\partial f}{\partial y_\beta}, \tag{57}$$

but it is a great improvement on Theorem 3, insofar as it does not involve a boundary condition on the *derivatives* of f.

THEOREM 8. *If the saltus of $f(\zeta)$ vanishes on C, then F is analytic in $z_1, \cdots, z_k$, that is, $\dfrac{\partial F}{\partial \bar{z}_\beta} = 0, \beta = 1, \cdots, k.$*

Also, in D_∞ the functional element $F_\infty(z)$ is identically 0.

For use in the proof, we introduce the symbol $s^{\bar{\alpha}}$ which is the differential form arising from

$$d\bar{\zeta}_1 d\zeta_1 \cdots d\bar{\zeta}_k d\zeta_k \tag{58}$$

by dropping $d\bar{\zeta}_\alpha$ and by $s^{\bar{\alpha}\bar{\beta}}$ we denote the symbol arising from (28) by dropping both $d\bar{\zeta}_\alpha$ and $d\bar{\zeta}_\beta$ if $\alpha \neq \beta$, and the symbol 0 if $\alpha = \beta$. Now

$$\frac{(2\pi i)^k}{(k-1)!}\frac{\partial F}{\partial \bar{z}_\beta} = \int_B - \frac{1}{(k-1)!}\sum_{\alpha=1}^{k} f\,\frac{\partial^2 V}{\partial \zeta_\alpha \partial \bar{\zeta}_\beta}\,d\zeta_\alpha\,\omega^{k-1}, \tag{59}$$

but the integral is, identically in f, V,

$$f\cdot \Delta V \cdot s^{\bar{\beta}} + \sum_{\alpha=1}^{k} d\left(f\frac{\partial V}{\partial \zeta_\alpha}\right)s^{\bar{\alpha}\bar{\beta}} - \sum_{\alpha=1}^{k}\frac{\partial V}{\partial \zeta_\alpha}\,df s^{\bar{\alpha}\bar{\beta}}. \tag{60}$$

In our case we have $\Delta V = 0$, and we also have $df.\ s^{\bar{\alpha}\bar{\beta}} = 0$ since f is independent of both $\bar{\zeta}_\alpha$ and $\bar{\zeta}_\beta$. Therefore, by Stokes' theorem, (59) is

$$\int_C f\,\frac{\partial V}{\partial \zeta_\alpha}\,s^{\bar{\alpha}\bar{\beta}}$$

and thus (59) vanishes if the saltus of f vanishes on C. The second half of the theorem has been proven in section 6.

Finally we prove the following analogue of Theorem 4.

THEOREM 9. *If the conglomerate function $f(\zeta_1, \cdots, \zeta_k)$ is analytic in B and if its saltus vanishes on C, then the following propositions hold:*

(i) *If the component D_∞ borders on a simplex B_p from both sides then $f_p(\zeta) = 0$.*

(ii) *If, in particular, $\tilde{B}$ does not decompose E_n, then $f(\zeta) = 0$.*

(iii) *If a simplex B_p separates D_∞ from D_τ then $f_p(\zeta)$ can be continued into all of D_τ.*

As a special conclusion we observe that if the set of all zeroes of an analytic function $f(\zeta_1, \cdots, \zeta_k)$ contains a bounding cycle of dimension $2k - 2$ then the function vanishes identically.

As a counterpart to this property we observe that if $f(\zeta_1, \cdots, \zeta_k)$ is $\neq 0$ everywhere on the boundary $\bar{B}$ of a domain D then $f \neq 0$ everywhere in D. In fact f^{-1} is analytic in a neighborhood of B and therefore it can be continued into all of D.

Chapter IV

Integrability Conditions

10. We are interested in the system of equations

$$(61)\qquad \begin{aligned} &\frac{\partial f}{\partial \bar{z}_\alpha} = f_\alpha(z, \bar{z}), && \alpha = 1, \cdots, k, \\ &\frac{\partial f_\alpha}{\partial \bar{z}_\beta} = \frac{\partial f_\beta}{\partial \bar{z}_\alpha}, && \alpha, \beta = 1, \cdots, k, \end{aligned}$$

in any domain D in E_{2k}. The functions f_α are given and the function f is a desired solution. The system is in some respects similar to the classical system

$$(62)\qquad \begin{aligned} &\frac{\partial f}{\partial x_\alpha} = f_\alpha(x), && \alpha = 1, \cdots, n, \\ &\frac{\partial f_\alpha}{\partial x_\beta} = \frac{\partial f_\beta}{\partial x_\alpha}, && \alpha, \beta = 1, \cdots, n, \end{aligned}$$

in a domain D in E_n. Assuming all functions f_α to be analytic in x or $(z, \bar{z})$ respectively then either system has an analytic solution in a neighborhood of a point. For (62) this follows easily if all $f_\alpha(x)$ are expanded in power series. Also if the functions $f_\alpha(x)$ contains additional analytic parameters then so does the solution f. Hence we may conclude the existence of local solutions of (61) by viewing $\bar{z}_1, \cdots, \bar{z}_k$ as independent variables and $z_1, \cdots, z_k$ as parameters.

As for solutions in the large, we will first prove a theorem which will apply to either system (61) or (62). The theorem will not be needed, but the proof will be based on an application of Green's formula which apparently has not been noticed before. After that we will prove another theorem for the system (61). It will be the exact counterpart within the scope of the general Green's integral to Cousin's method as based on the classical Cauchy's integral.

Theorem 10. *If D is a simplicial domain with boundary $\bar{B}$, if $\{f_\alpha\}$ is given in D and if either system* (61) *or* (62) *has a solution in $\bar{B}$, then there also exists a solution in D.*

We will treat only system (62) for system (61) the proof is analogous. With the given functions f_α in $\bar{D}$ and the solution f in $\bar{B}$ we set up the integral

$$\gamma_n F(x) = \int_B f \frac{\partial G}{\partial \xi_\alpha} \sigma^\alpha - \int_D f_\alpha \frac{\partial G}{\partial \xi_\alpha} dV_\xi .$$

By formula (15), if f exists in all of $\tilde{D}$, then $F(x) = f(x)$. Now take a fixed point x^0 in D and a sufficiently small neighborhood D' with boundary $\tilde{B}'$. If it is sufficiently small, then there exists a function $g(x)$ with $f_\alpha = \frac{\partial g}{\partial x_\alpha}$ in D' such that, for x in D',

$$\gamma_n g(x) = \int_{B'} f \frac{\partial G}{\partial \xi_\alpha} \sigma^\alpha - \int_{D'} f_\alpha \frac{\partial G}{\partial \xi_\alpha} dV_\xi .$$

The reader will now easily see that our theorem will follow from the following

LEMMA. *If the boundary B of a domain D consists of two disjoint parts, $B = B' + B''$, if f in $\tilde{B}'$ and g in $\tilde{B}''$ are each a solution of* (62), *then for x in $E_n - \bar{D}$, the partial derivatives of the function*

$$\int_{B'} f \frac{\partial G}{\partial \xi_\alpha} \sigma^\alpha + \int_{B''} g \frac{\partial G}{\partial \xi_\alpha} \sigma^\alpha - \int_D f_\alpha \frac{\partial G}{\partial \xi_\alpha} dV_\xi$$

vanish identically.

In order to prove the lemma we note that the partial derivative of the given expression with respect to x_β, for x outside $\bar{D}$, is

$$-\int_{B'} f \frac{\partial^2 G}{\partial \xi_\alpha \partial \xi_\beta} \sigma^\alpha - \int_{B''} g \frac{\partial^2 G}{\partial \xi_\alpha \partial \xi_\beta} \sigma^\alpha + \int_D f_\alpha \frac{\partial^2 G}{\partial \xi_\alpha \partial \xi_\beta} dV_\xi .$$

By (31) this is

$$\int_B f_\beta \frac{\partial G}{\partial \xi_\alpha} \sigma^\alpha - \int_B f_\alpha \frac{\partial G}{\partial \xi_\alpha} \sigma^\beta + \int_D f_\alpha \frac{\partial^2 G}{\partial \xi_\alpha \partial \xi_\beta} dV_\xi .$$

On applying Stokes' theorem to the integrals over B, this turns out to be

$$\int_D \left(\frac{\partial f_\beta}{\partial \xi_\alpha} - \frac{\partial f_\alpha}{\partial \xi_\beta}\right) \frac{\partial G}{\partial \xi_\alpha} dV_\xi$$

and this in its turn is obviously 0.

11. An existence theorem. We will now prove a genuine existence theorem.

THEOREM 11. *The system* (61) *has a solution if the domain D is a general polycylinder, that is the topological product of a domain in each of the variables $z_1, \cdots, z_k$ separately.*

The theorem will follow from a lemma on functions in the plane of one complex variable $z = x + iy$.

If $\varphi(z, \bar{z})$ is analytic in $\tilde{D}$, then the integral

$$F(z, z) = \frac{-1}{2\pi i} \int_D \frac{\varphi(\zeta, \bar{\zeta})\, d\bar{\zeta}\, d\zeta}{\zeta - z} \tag{63}$$

is analytic in z, $\bar{z}$ and $\frac{\partial F}{\partial \bar{z}} = \varphi(z, \bar{z})$. Also if φ depends analytically on additional parameters, then so does F.

For z outside $\bar{D}$ the integral (63) is obviously analytic in z and other parameters and independent of $\bar{z}$. Hence, in order to prove that the stated properties of $F(z, z)$ hold in the neighborhood of a point z^0 in D we may assume that D is an arbitrarily small neighborhood of z^0. If the neighborhood is sufficiently small then there exists a function $\Phi(z, \bar{z})$ of which $\varphi(z, \bar{z})$ is the derivative with respect to $\bar{z}$, and thus we have to prove the properties of $F(z, \bar{z})$ only in the specialized case

$$F(z, \bar{z}) = \frac{-1}{2\pi i}\int_D \frac{\partial\Phi}{\partial\bar{\zeta}}\frac{d\bar{\zeta}\, d\zeta}{\zeta - z}. \tag{64}$$

If we introduce the Newtonian kernel

$$V = \log|\zeta - z|^2,$$

then (64) is

$$F(z, \bar{z}) = \frac{-1}{2\pi i}\int_D \frac{\partial\Phi}{\partial\bar{\zeta}}\frac{\partial V}{\partial\zeta}\, d\bar{\zeta}\, d\zeta. \tag{65}$$

Now, formula (48) which we have claimed only for $k \geqq 2$ also holds for $k = 1$. Thus, for z in D, we have

$$\Phi(z, \bar{z}) = \frac{1}{2\pi i}\int_B \Phi\,\frac{d\zeta}{\zeta - z} + F(z, \bar{z}). \tag{66}$$

But the boundary integral is independent of $\bar{z}$, and this proves our lemma.

We are now going to prove Theorem 11. Since we are going to apply induction on k, we will have to prove our theorem in the sharper form that some solution of (61) depends analytically on any parameters occurring in f_β. We assume the existence of a solution of

$$\frac{\partial f}{\partial\bar{z}_\beta} = f_\beta(z_1, z_1; z_2, z_2, \cdots, z_k, z_k) \tag{67}$$

in $\beta = 2, \cdots, k$; with $z_1, \bar{z}_1$ as parameters. Denoting the solution of (67) by h and introducing the new function $g = f - h$, then the system (61) reduces to a system

$$\frac{\partial g}{\partial\bar{z}_1} = g_1(\bar{z}_1; z_1, z_2, \cdots, z_k)$$

$$\frac{\partial g}{\partial\bar{z}_2} = \cdots = \frac{\partial g}{\partial\bar{z}_k} = 0,$$

and this system is solvable by our lemma.

Chapter V

Meromorphic Functions in Euclidean Space

12. A simplified problem of Mittag-Leffler type. We are now returning to the topic of Chapter II, and the symbols $C_1, \cdots, C_r; B_1, \cdots, B_s; D_1, \cdots, D_t$,

D_∞ shall have the previous meaning. In addition to the incidence number ϵ_{pq} between C_p and B_q we also introduce the incidence number $\eta_{q\tau}$ between B_q and D_τ. An elementary combinatorial statement ("the boundary of a boundary is zero") implies the relation

$$\sum_{q=1}^{s} \epsilon_{pq}\eta_{q\tau} = 0 \tag{68}$$

and its validity is not impaired by the fact that the domains D_τ are not necessarily simplices. We furthermore introduce an arbitrary pointset S having the following properties: it is closed and nowhere dense, and if R is any domain then $R - S$ (that is $R - R\cdot S$) is also a domain. The set S will be termed an "exceptional" set. Obviously if a harmonic function φ in $R - S$ has an analytic continuation in all of R then the continuation is unique. If such a continuation exists we will say that "the singularities of φ at S are removable" or that "φ exists in (all of) R," or that "φ is analytic in R."

We now assume the existence of a harmonic function φ_τ in $\tilde{D}_\tau - S$, for each $\tau = 1, \cdots, t, \infty$, and we further assume that for each B_q the saltus

$$f_q = \sum_{\tau=1}^{\infty} \eta_{q\tau}\varphi_\tau \tag{69}$$

is analytic not only in $\tilde{B}_q - S$ but in all of $\tilde{B}_q$. We therefore may introduce the conglomerate function

$$F(x) = \frac{1}{\gamma_n}\sum_{q=1}^{s}\int_{B_q}\left(f_q\frac{\partial G}{\partial \xi_\alpha} - G\frac{\partial f_q}{\partial \xi_\alpha}\right)\sigma^\alpha. \tag{70}$$

Also, the saltus of $\{f_q\}$ in $\tilde{C}_p$ is defined as

$$f^p = \sum_q \epsilon_{pq}f_q = \sum_q\sum_\tau \epsilon_{pq}\eta_{q\tau}\varphi_\tau$$

and owing to relation (68) it turns out to be identically 0. Thus we may apply Theorems 1, 2, 3 and we obtain the following result.

THEOREM 12. *If the conglomerate function* $\{\varphi_\tau\}$ *is defined and harmonic in* $\{\tilde{D}_\tau - S\}$, *and if its saltus* $\{f_p\}$ *has none other than removable singularities in* $\tilde{B}$, *then the same saltus will be produced by the harmonic conglomerate function* (70) *which has no singularities whatsoever in* $\{\tilde{D}_\tau\}$. *Also, if* $\{\varphi_\tau\}$ *is a solution of a system of differential equations with constant coefficients,*

$$\Lambda\varphi = 0, \tag{71}$$

then $\{F_\tau\}$ *is also a solution.*

Thus, the functional elements

$$F_1 - \varphi_1 = F_2 - \varphi_2 = \cdots = F_\infty - \varphi_\infty$$

define a function Φ *with the following properties. It is one functional element in* $E_n - S$, *it is harmonic and a solution of* (71), *and in each* $\tilde{D}_\tau$, *the difference* $\Phi - \varphi_\tau$ *is analytic without exception.*

A seemingly striking situation arises, if the system (71) has the uniqueness property, and the set S is bounded. Being analytic outside a bounded domain, the function Φ must be analytic in the entire space and therefore $\varphi_\tau \equiv \Phi - F_\tau$ has only removable singularities in $\bar{D}_\tau$ each. However, it should be noted that in this case, to start with, φ_∞ itself has to have an analytic continuation in the entire space. Now, let B'_λ be one of our faces lying in the boundary of D_∞. If it separates D_∞ from D_∞ then the saltus of φ is $\varphi_\infty - \varphi_\infty \equiv 0$, and thus naturally has no singularities. If however, B'_λ separates D_∞ from another domain D_c, then the saltus is $\pm(\varphi_\infty - \varphi_c)$ and if it is assumed to be free from singularities then so is φ_c itself. In other words: if we introduce the closet set $\bar{D}_1 + \cdots + \bar{D}_t$, and if we consider its interior D_0 and the frontier B_0; then our assumption requires *explicitly* that in a neighborhood $\tilde{B}_0$ the function $\{\varphi_\tau\}$ itself shall have no singularities, and only for "internal" faces B''_μ is the explicit requirement less stringent. If such a face separates two domains D_a, D_b then the assumption requires only that the saltus $\varphi_a - \varphi_b$ shall have no singularities. The theorem then allows the conclusion that actually the singularities of each φ_τ itself are removable everywhere in $\bar{D}_\tau$.

13. A simplified problem of Weierstrass type. Everything else being as before, we consider a conglomerate function $\{\psi_\tau(x)\}$ with the following properties. Each $\psi_\tau(x)$ is analytic in $\tilde{D}_\tau$ and different from 0 in $\tilde{D}_\tau - S$. The "multiplicative" saltus

$$\prod_{\tau=1}^{\infty} \psi_\tau^{\eta_{q\tau}}$$

is by assumption analytic everywhere in $\tilde{B}_q$, and in the neighborhood of every point of $\tilde{D}_\tau - S$ the function $\log \psi_\tau(x)$ and its partial derivatives of first order both harmonic and a solution of (71), the latter having uniqueness property.

For fixed α, we introduce the conglomerate function

$$\varphi_\tau(x) = \frac{\partial}{\partial x_\alpha} \log \psi_\tau(x) = \frac{1}{\psi_\tau} \frac{\partial}{\partial x_\alpha} \psi_\tau,$$

and by Theorem 12 it is analytic in $\tilde{D}_\tau$. Thus, for fixed τ there exist analytic functions $F_1(x), \cdots, F_n(x)$ in $\tilde{D}_\tau$ such that

$$\frac{\partial \psi}{\partial x_\alpha} = \psi \cdot F_\alpha, \qquad \alpha = 1, \cdots, n.$$

On expanding ψ in a power series in a neighborhood of any point in $\tilde{D}_\tau$, the reader will easily convince himself that *every $\psi_\tau(x)$ is different from 0 in all of $\tilde{D}_\tau$*.

14. The theorem of Cousin. Returning to the topic of section 12, we now set up the modified integral

$$F^0(x) = \frac{1}{\gamma_n} \sum_{\mu=1}^{m} \int_{B''_\mu} \left(f_\mu \frac{\partial G}{\partial \xi_\alpha} - G \frac{\partial f_\mu}{\partial \xi_\alpha} \right) \sigma^\alpha$$

extended only over the "internal" faces B''_μ. Our present setup implies that we require the saltus function f_q to be analytic throughout $\tilde{B}_q$ only in case of internal faces, with no assumption made about the other faces.

Theorems 1 and 2 will permit the conclusion that for each D_a *whose closure is contained in* D_0 the corresponding functional element F_a^0 has no singularities in $\tilde{D}_a$, and if two such simplices have a face B''_μ in common, then again

$$F_a^0 - \varphi_a = F_b^0 - \varphi_b. \tag{72}$$

Therefore if a number of such simplices $D_a, D_b, \cdots$ are the simplicial subdivision of a subdomain D^0 of D_0, then the functional elements (72) merge into a function Φ^0 in $\tilde{D}^0 - S$. However it must not be assumed that Φ^0 will be a solution of (71) because $\{\varphi_r\}$ is. All that can be said about the function (or functions)

$$h^0 = \Lambda\Phi^0 = \Lambda F^0 - \Lambda\varphi = \Lambda F^0$$

is that it is harmonic without singularities in D^0 (and indeed in $\tilde{D}^0$). We may now pose the problem of constructing a harmonic solution of the "unhomogeneous" system

$$\Lambda H = h^0$$

for any harmonic h^0 in any (or some specialized type of) domain D^0. If such a solution exists then $\Phi^0 - H$ will have in D^0 the same properties as were stated in Theorem 12 for the function itself.

A classical case arises if E_n is the space E_{2k} of the variables $z_\alpha = x_\alpha + iy_\alpha$, $\alpha = 1, \cdots, k$. Each φ_r shall be a mereomorphic element in $\tilde{D}_r$, that is a quotient of two analytic functions in $z_1, \cdots, z_k$ alone; with S being the union of the zero-sets of the denominators. The system (71) is

$$\frac{\partial\varphi}{\partial\bar{z}_1} = 0, \cdots, \frac{\partial\varphi}{\partial\bar{z}_k} = 0. \tag{73}$$

Theorem 11 will now give in substance a new proof for the classical theorem of Cousin about merging meromorphic elements in a polycylinder into one meromorphic function.

CHAPTER VI

Meromorphic Functions on Torus Space

15. Green's formula. We consider the locally Euclidean n-dimensional torus T_n with all periods 1:

$$T_n: \quad 0 \leqq x_\alpha < 1, \qquad \alpha = 1, \cdots, n.$$

Stokes' theorem (1) for ordinary or conglomerate functions remains in force. The scalar φ has to be strictly a function on the torus; that is, as a function in E_n it has to be strictly periodic in each variable x_α with period 1. However, any quantity ψ appearing only in a differential $d\psi$ is permitted to alter by an

additive constant if continued along a closed path in T_n. We now consider on the torus a function $G(\xi, x) = G(\xi - x)$ of the type (16). Formulas (14) and (15) are valid again, and if we apply them to the special domain $D = T_n$, $B = 0$, we obtain

$$\gamma_n f(x) = -\int_{T_n} \left(f \cdot \Delta G + \frac{\partial f}{\partial \xi_\alpha} \frac{\partial G}{\partial \xi_\alpha} \right) dV_\xi \tag{74}$$

$$\gamma_n f(x) = -\int_{T_n} (f \cdot \Delta G - G \cdot \Delta f) \, dV_\xi \tag{75}$$

On putting $f(\xi) = 1$ we obtain in particular

$$\gamma_n = -\int_{T_n} \Delta G \cdot dV_\xi ,$$

thus ΔG cannot be 0. But it might be a constant, and if so, the constant must of necessity be $-\gamma_n$. That is,

$$\Delta G = -\gamma_n . \tag{76}$$

Furthermore a solution of (76) must of necessity be unique. The actual existence of such a G can be proven in many ways. Suffice it to state here, that except for a constant factor, it is the sum of the series

$$\sum_{h_1, \cdots, h_n = -\infty}^{\infty}{}' \frac{\exp \{2\pi i [h_1(x_1 - \xi_1) + \cdots + h_n(x_n - \xi_n)]\}}{h_1^2 + \cdots + h_n^2} .$$

The series is not absolutely convergent but can be "summed" by general summation processes. From now on, G will be this solution of (76). Therefore, we now have

$$\begin{matrix} \gamma_n f(x) \\ 0 \end{matrix} = \int_B \left(f \frac{\partial G}{\partial \xi_\alpha} - G \frac{\partial f}{\partial \xi_\alpha} \right) \sigma^\alpha + \gamma_n \int_D f \, dV_\xi \tag{77}$$

and in particular

$$\begin{matrix} \gamma_n f(x) \\ 0 \end{matrix} = \int_B \left(f \frac{\partial G}{\partial \xi_\alpha} - G \frac{\partial f}{\partial \xi_\alpha} \right) \sigma^\alpha + \textit{constant}. \tag{78}$$

16. Green's integral. We again consider simplices $B_1, \cdots, B_s; C_1, \cdots, C_r$. The complement $T_n - \bar{B}$ decomposes into domains $D_1, \cdots, D_t$, with no "D_∞" being present. We again set up the integral

$$F(x) = \frac{1}{\gamma_n} \int_B \left(f \frac{\partial G}{\partial \xi_\alpha} - G \frac{\partial f}{\partial \xi_\alpha} \right) \sigma^\alpha \tag{79}$$

for an arbitrary harmonic conglomerate function $f(\xi)$ in $\tilde{B}$, and we note that Theorems 1 and 2 carry over literally. In the proof to Theorem 1 relation (77)

has to be used instead of (20). In the proof to Theorem 2, the value of (25) is not 0, but some other constant, namely

$$-\gamma_n \sum_{q=0}^{\mu} \int_{B_q} f_q \, dV_\xi$$

and thus, in the proof, $\gamma_n(F(x) - F'(x))$ is constant, and hence $F(x)$ continuable.

The remarkable thing about the torus is the way in which Theorems 1 and 2 merge immediately with a part of Theorem 12.

THEOREM 13. *If $f(\xi)$ is a harmonic conglomerate function in $\tilde{B}$ and if its saltus vanishes identically in $\tilde{C}$, then f is in its turn the saltus of the harmonic conglomerate function F in $D \equiv T_n$.*

In particular, if S is an exceptional set, if φ is a conglomerate harmonic function in $\tilde{D} - S$, and if its saltus f has no singularities in $\tilde{B}$, then the function F is a harmonic conglomerate function without singularities having the same saltus as φ.

Finally we will analyze Theorem 3.

17. Meromorphic functions. Denoting again (79) by Lf we see on the basis of (31)–(34) that

$$L \frac{\partial f}{\partial \xi_\beta} - \frac{\partial}{\partial x_\beta} Lf = -\gamma_n \int_B f\sigma^\beta \tag{80}$$

and that in particular we have

$$L \frac{\partial f}{\partial \xi_\beta} - \frac{\partial}{\partial x_\beta} Lf = \textit{constant}; \tag{81}$$

provided again that the saltuses of $f, \dfrac{\partial f}{\partial \xi_1}, \cdots, \dfrac{\partial f}{\partial \xi_n}$ vanish on C. Thus we will have

$$L \frac{\partial f}{\partial \xi_\beta} - \frac{\partial}{\partial x_\beta} Lf = 0$$

if and only if

$$\int_B f\sigma^\beta = 0. \tag{82}$$

It will now be easy for the reader to verify the following analogue to Theorem 3.

THEOREM 14. *If $f(\xi)$ is harmonic in $\tilde{B}$ and if its saltus vanishes identically in $\tilde{C}$, then*

$$L\Lambda f - \Lambda Lf = \textit{constant} \tag{83}$$

If Λ be represented in some (or any) way as

$$\Lambda \equiv \sum_{\beta=1}^{n} \frac{\partial}{\partial x_\beta} \Lambda_\beta \tag{84}$$

where $\Lambda_1, \cdots, \Lambda_n$ are any operators, then

$$L\Lambda f - \Lambda Lf = -\gamma_n \int_B \left(\sum_{\beta=1}^{n} \Lambda_\beta f \cdot \sigma^\beta\right). \tag{85}$$

In particular we have

$$L\Lambda f = \Lambda Lf$$

if and only if

$$\int_B \left(\sum_{\beta=1}^{n} \Lambda_\beta f \cdot \sigma^\beta\right) = 0. \tag{86}$$

Therefore, if $\Lambda f = 0$ *and if* (86) *holds, then* ΛF *is likewise* 0.

For

$$\Lambda f = \left(\frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_m^2}\right) f, \qquad m < n$$

(86) is

$$\int_B \left(\sum_{\beta=1}^{n} \frac{\partial f}{\partial x_\beta} \sigma^\beta\right) = 0$$

Or, if

$$\Lambda f = \frac{\partial f}{\partial x_1} + i \frac{\partial f}{\partial x_2} = 2 \frac{\partial f}{\partial \bar{z}},$$

where

$$z = x_1 + ix_2,$$

then we may put $\Lambda_1 = 1, \Lambda_2 = i, \Lambda_3 = \cdots = \Lambda_n = 0$, and therefore

$$\sum_{\beta=1}^{n} \Lambda_\beta f \cdot \sigma^\beta = f \cdot (d\xi_1 - id\xi_2) \cdot \sigma^{1,2} \equiv f\, d\bar{\xi} \cdot \sigma^{1,2}.$$

Turning now to the torus space T_{2k} of the complex variables $z_\alpha = x_\alpha + iy_\alpha$, $\alpha = 1, \cdots, k$, we therefore obtain the following analogue to Theorem 12.

THEOREM 15. *If* $\varphi_\tau(z_1, \cdots, z_k)$ *is meromorphic in* $\tilde{D}_\tau$, $\tau = 1, \cdots, t$, *but each* $\varphi_a - \varphi_b$ *is analytic in* $\tilde{B}_q$ *and if the k periodicity conditions*

$$\int_B f\, d\bar{\xi}_\alpha \cdot \omega^{k-1} = 0, \qquad \alpha = 1, \cdots, k \tag{87}$$

are fulfilled, then there exists a meromorphic function $F(z_1, \cdots, z_k)$ *on the torus, such that* $F - \varphi_\tau$ *is analytic in* $\tilde{D}_\tau$.

If the conditions (87) are not stipulated, then by (83) we have at any rate

$$\frac{\partial F}{\partial \bar{z}_\alpha} = c_\alpha, \qquad \alpha = 1, \cdots, k$$

where c_α is a constant. In this case the function

$$F(z_1, \cdots, z_k; \bar{z}_1, \cdots, \bar{z}_k) - c_1\bar{z}_1 - \cdots - c_k\bar{z}_k$$

in the plane will be strictly meromorphic, however if continued along a closed path on the torus it alters by an additive constant which need not be 0.

The additive constants do not alter the algebraic singularities, and in this sense we have

THEOREM 16. *If an exceptional set S on T_{2k} is locally the singular set of a meromorphic function, then it has the same property in the large.*

PRINCETON UNIVERSITY

ANNALS OF MATHEMATICS
Vol. 44, No. 4, October, 1943

DISCONTINUOUS GROUPS

BY CARL LUDWIG SIEGEL

(Received May 5, 1943)

I. INTRODUCTION

1. Let G be a topological group, and let H be a *discrete* subgroup of G; this means that there exists in G a neighborhood U of the unit element e such that no other element of H is contained in U. If M is a subset of G and $a \in H$, then the set Ma is an *image* of M. A *fundamental set* F relative to H is defined by the following three properties: 1) $FH = G$; 2) $Fa \cap F = 0$, whenever $e \neq a \in H$; 3) F is a Borel set in G. The first two properties establish that every point x of G is covered by one and only one image of F. We obtain an arbitrary set M with these two properties, if we select a representative x from every left coset xH of H relative to G; however, in general, such a set M is not a Borel set. In Lemma 2, we shall prove that a fundamental set exists, if G satisfies the second axiom of countability.

We say that a fundamental set F is *normal*, if every point of G has a neighborhood contained in the union of a finite number of images of F. An image Fa of F is called a *neighbor* of F, if $\bar{F}a \cap \bar{F} \neq 0$. Let H_0 be the group generated by all elements $a \in H$ satisfying $\bar{F}a \cap \bar{F} \neq 0$. In section 9 we shall prove, for any connected group G satisfying the second axiom of countability, that $H_0 = H$, whenever the fundamental set F is normal; in particular, if F has only a finite number of neighbors, then H possesses a finite system of generators.

2. Let T be a topological space of points τ, and let $\tau \rightarrow f(\tau, x)$, $x \in G$, be an open continuous *representation* of G as a transitive group of homeomorphic mappings of T on itself. The representation $\tau \rightarrow f(\tau, a)$, $a \in H$, of H in T is called *discontinuous*, if no sequence $f(\tau, a_n)$ $(n = 1, 2, \cdots)$ converges to a point in T, as a_n runs over distinct elements of H. Let τ_1 be a given point of T, and consider the set C of elements $x \in G$ satisfying $f(\tau_1, x) = \tau_1$; obviously C is a closed subgroup of G. It is known that then T and $C\backslash G$ are homeomorphic, where the topological space $C\backslash G$ consists of the right cosets Cy of C relative to G; on the other hand, for any closed subgroup C of G and any homeomorphic mapping of $C\backslash G$ on a topological space T, the transformations $Cy \rightarrow Cyx$, $x \in G$, define an open continuous representation of G as a transitive group of homeomorphic mappings of T on itself. Therefore, in order to find all discontinuous representations of H, we can restrict ourselves to the investigation of the case $T = C\backslash G$, where C is any closed subgroup of G. In Lemma 6 we shall prove that the representation of H in $C\backslash G$ is discontinuous in the case of a compact group C.

Let G be a locally compact group satisfying the second axiom of countability. For all Borel sets B in G, the Haar measure defines a completely additive and right-invariant volume $v(B)$, which is positive for all open sets. It will be proved that the volume of a fundamental set F does not depend upon the parti-

cular choice of F and only upon the discrete subgroup H. The main result of the second chapter is the following theorem.

THEOREM 1: *Let $v(F)$ be finite; then the representation of H in $C\backslash G$ is discontinuous, if and only if the closed subgroup C is compact.*

The interest of Theorem 1 lies in the *necessity* of the condition concerning C. The proof uses an idea similar to that in the proof of Poincaré's theorem in ergodic theory.

3. Let the subgroup C be compact, and assume that the representation of H in $C\backslash G$ has no fixed points; this means that $Cya \neq Cy$ for all $y \in G$ and all $a \neq e$ in H. It can readily be seen that this assumption is fulfilled, in particular, if H does not contain any element of finite order except e. We shall prove, in Lemma 7, that there exists a fundamental set F relative to H such that $F = CF$, whenever G satisfies the second axiom of countability; then F may be considered as a fundamental set in the space $C\backslash G$.

Let G be a locally compact group satisfying the second axiom of countability. We say that H is a subgroup of the *first kind* in G, if there exists a normal fundamental set relative to H, of finite volume, having only a finite number of neighbors. We have no general constructive method of deciding whether a given subgroup H is of the first kind; however, in the known particular cases where we are able to answer this question, the problem is simplified by considering it for the space $C\backslash G$ instead of G, with suitably chosen compact C. In the case of the *unimodular* group, e.g., $G = G_m$ is the multiplicative group of all real m-rowed matrices x with determinant ± 1, and $H = H_m$ is the subgroup consisting of all $x = u$ with integral elements; if C denotes the orthogonal group in m dimensions, then the space $C\backslash G$ is homeomorphic to the space T of the positive symmetric matrices $t = x'x$ with determinant 1, and H_m is represented by the transformations $t \rightarrow u'tu$ of T into itself. In this way, Minkowski replaced the problem of constructing a fundamental set for H_m in G_m by the corresponding problem concerning the reduction of positive quadratic forms in m variables, and he proved that H_m is a subgroup of the first kind in G_m.

The problem of finding *all* subgroups H of the first kind in G seems to be far beyond our power, even if we consider only the particular case $G = G_m (m \geqq 2)$. The solution is known only for $m = 2$:

G_2 is the group of all two-rowed matrices $\begin{pmatrix} ab \\ cd \end{pmatrix}$ with real elements and the determinant 1. A subgroup H of G_2 is of the first kind, if and only if the representation of H by the linear mappings $z \rightarrow (az + b)(cz + d)^{-1}$ possesses in the upper z half-plane a fundamental polygon with a *finite* number of vertices. However, even in this comparatively simple case, we have no method of deciding whether two arbitrarily given subgroups H and H_0 of G_2 are isomorphic.

4. In the third chapter, we investigate the properties of a special type of discrete groups; these groups include the unimodular group H_m, and they are found by the following considerations:

We assume that the topological group G may be imbedded in a connected topological ring A which is locally compact and satisfies the second axiom of countability. It follows from a theorem of Jacobson and Taussky[1] that A is isomorphic to an algebra of finite rank in R, the field of real numbers. Consequently, we can introduce the norm $N(x)$ of an arbitrary element x in A. Let G be the group of all x with $N(x) = \pm 1$.

On the other hand, let A_0 be a simple algebra of finite rank in R_0, the field of rational numbers. Under extension of R_0 into R, the simple algebra A_0 in R_0 is replaced by a semi-simple algebra in R, and we assume now that this is the algebra A. Let an *order* J in A_0 be given; a quantity a is called a *unit* in J, if both a and a^{-1} belong to J. It is obvious that the units in J constitute a discrete subgroup H in G.

THEOREM 2: *The group of units in a simple order is of the first kind.*

The proof of this theorem depends upon the theory of reduction of positive quadratic forms; it is a generalization of Minkowski's proof concerning the unimodular group H_m.

5. By a well-known theorem of Wedderburn, the simple algebra A_0 is isomorphic to an algebra of matrices whose elements are arbitrary quantities in a division algebra D_0. Let Z be the center of D_0. In two important special cases the theory of the group of units in J had been investigated a long time since, namely in the case $A_0 = Z$ by Dirichlet, and in the case $D_0 = R_0$ by Minkowski. Recently, Humbert[2] studied the more general case $D_0 = Z$, and Weyl[3] proved Humbert's results as simple consequences of a geometric theory of reduction of lattices, which he applied also in the case of a totally definite quaternion algebra D_0 over a totally real center Z.

Concerning the general case of the group H of units in an arbitrary simple order, Eichler[4] stated that H has a finite system of generators; however, his proof is correct only for $A_0 = D_0$. Moreover, Eichler's paper contains an interesting theorem about continuous mappings of G/H on manifolds with the Poincaré group H, also without complete proof; it is related to Theorem 1, and it was the starting-point of the researches in the second chapter.

6. Let an *involution* in A_0 be given, i.e., a mapping $x \rightarrow x^*$ of A_0 onto itself such that $(x^*)^* = x$, $(x + y)^* = x^* + y^*$, $(xy)^* = y^* x^*$; under extension of R_0 into R, we obtain an involution in A. Let $s \in A_0$, $N(s) \neq 0$ and $s^* = s$, and consider the set $G(s)$ of all $x \in A$ such that $x^*sx = s$; obviously $G(s)$ is a subgroup

[1] N. Jacobson and O. Taussky, *Locally compact rings*, Proc. Nat. Acad. Sci. 21, pp. 106–108 (1935).

[2] P. Humbert, *Théorie de la réduction des formes quadratiques définies positives dans un corps algébrique K fini*, Comment. Math. Helv. 12, pp. 263–306 (1940).

[3] H. Weyl, *Theory of reduction for arithmetical equivalence. II*, Trans. Amer. Math. Soc. 51, pp. 203–231 (1942).

[4] M. Eichler, *Zur Einheitentheorie der einfachen Algebren*, Comment. Math. Helv. 11, pp. 253–272 (1939).

of G. For any given order J in A_0, the intersection $H \cap G(s)$ defines a discrete subgroup $H(s)$ of $G(s)$.

Assume now that the involution $x \to x^*$ leaves invariant all elements of the center Z. By the results of Albert,[5] the division algebra D_0 is then either Z itself or a quaternion algebra over Z.

In the first case, A_0 is the ring of all m-rowed matrices x with elements in Z, and x^* is the transpose of x; then $H(s)$ is the group of units of the symmetric matrix s. In particular, let $Z = R_0$; then it is known[6] that $H(s)$ is a subgroup of the first kind in $G(s)$, except in the trivial case $m = 2$, $N(s) = -r^2$, r a rational number. This result was proved by an application of Hermite's method of continuous reduction of indefinite quadratic forms: We consider the space T of all positive symmetric matrices t, with real elements, satisfying $s = ts^{-1}t$; the transformations $t \to x^*tx$, $x \in G(s)$, are transitive in T, and the group C of all x satisfying $t = x^*tx$ is compact, for any fixed t; from the theory of reduction of positive quadratic forms it follows that the representation $t \to a^*ta$, $a \in H(s)$, is of the first kind in T, except in the above mentioned trivial case. If $n, m - n$ is the signature of s, then T has $n(m - n)$ dimensions; it can easily be shown, by Theorem 1, that any discontinuous representation of $H(s)$ is at least $n(m - n)$-dimensional.

In the second case, A_0 is the ring of all m-rowed matrices x with elements in the quaternion algebra D_0 over Z, and x^* is the conjugate transpose of x. In particular, let Z be totally real, of degree h over R_0, and let the norm, relative to Z, of every element $\neq 0$ in D_0 be positive at $h - 1$ infinite prime spots of Z. It is known[7] that then $H(e)$ is of the first kind in $G(e)$; it can easily be proved, by Theorem 1, that any discontinuous representation of $H(e)$ is at least $m(m + 1)$-dimensional.

These examples indicate that a systematic investigation of all unit groups of fixed points in an involution might be of some interest.

II. General Theory

7. Throughout the present chapter, G is a topological group and H is a discrete subgroup of G.

Lemma 1: *For every $x \in G$ there exists a neighborhood V_x of x such that $V_x a \cap V_x = 0$, for all elements $a \neq e$ in H.*

Proof: Let U be a neighborhood of e containing no other element a of H. Since $e^{-1}e = e$, there exists a neighborhood V of x such that $V^{-1}V \subset U$. Define $V_x = xV$, and let y, z be two arbitrary points in V_x; then $y^{-1}z \in U$, hence $y^{-1}z \neq a$, $ya \neq z$, and V_x has the required property.

Lemma 1 establishes that the neighborhood V_x of x contains at most one element from every left coset yH of H relative to G.

---

[5] A. A. Albert, *Structure of algebras*, New York (1939); Chap. X.

[6] C. L. Siegel, *Einheiten quadratischer Formen*, Abh. Math. Sem. Hansischen Univ. 13, pp. 209–239 (1940).

[7] C. L. Siegel, *Symplectic geometry*, Amer. Jour. Math. 55, pp. 1–86 (1943).

LEMMA 2: *Let G satisfy the second axiom of countability; then there exists a fundamental set relative to H.*

PROOF: Let $A_1, A_2, \cdots$ constitute a basis of G. We consider all A_k contained in at least one of the neighborhoods V_x defined in Lemma 1, and we denote these A_k by $B_1, B_2, \cdots$. Let

$$F_1 = B_1, \qquad F_k = B_k - (B_1 \cup \cdots \cup B_{k-1})H \qquad (k = 2, 3, \cdots),$$
$$F = F_1 \cup F_2 \cup \cdots.$$

Since any V_x is the union of a certain number of sets B_k, every element x of G is contained in some B_k; in particular, this is true for every element a of H. On the other hand, by Lemma 1, B_k does not contain two elements of H. Consequently, H is finite or countably infinite, and F is a Borel set.

We have $F_k \subset B_k (k = 1, 2, \cdots)$; therefore, by Lemma 1,

$$F_k a \cap F_k = 0 \qquad (e \neq a \in H);$$

moreover

$$F_k a \subset (B_1 \cup \cdots \cup B_{l-1})H, \qquad F_k a \cap F_l = 0 \qquad (1 \leqq k < l;\ a \in H);$$

hence $F_k a \cap F_l = 0$, for arbitrary positive integers k, l and $e \neq a \in H$. It follows that $Fa \cap F = 0$, for $e \neq a \in H$.

Let x be any element of G. Since x lies in some B_k, there exists also an index k such that x lies in $B_k H$, but not in $(B_1 \cup \cdots \cup B_{k-1})H$; then $x \in B_k H - (B_1 \cup \cdots \cup B_{k-1})H = F_k H$, $x \in FH$. It follows that $FH = G$.

We have verified that F has the three characteristic properties of a fundamental set relative to H, and the lemma is proved.

If $S_1, S_2, \cdots$ is a finite or countably infinite number of disjoint sets, we denote their union by the sign $\sum$.

LEMMA 3: *Let G satisfy the second axiom of countability, and let E, F be two fundamental sets relative to H; then there exist two decompositions*

$$E = \sum_{a \in H} E_a, \qquad F = \sum_{a \in H} F_a,$$

where E_a and $F_a = E_a a$ are Borel sets.

PROOF: Define $Ea \cap F = F_a$, $E \cap Fa^{-1} = E_a$; then E_a and $F_a = E_a a$ are Borel sets; furthermore

$$F_a \cap F_b = Ea \cap Eb \cap F = 0 \qquad (a \in H; b \in H; a \neq b),$$
$$E_a \cap E_b = E \cap Fa^{-1} \cap Fb^{-1} = 0$$

and

$$\sum_{a \in H} F_a = (\sum_{a \in H} Ea) \cap F = (EH) \cap F = G \cap F = F,$$
$$\sum_{a \in H} E_a = E \cap (\sum_{a \in H} Fa^{-1}) = E \cap (FH) = E \cap G = E;$$

q.e.d.

8. Let us assume that there exists a fundamental set F, relative to H, such that $\bar{F}$ is compact, and consider any infinite set of cosets xH, $x \in S$. We choose $a_x \in H$ such that $xa_x \in F$; then the set xa_x, $x \in S$, has a limit point y in $\bar{F}$, hence the set of points xH, $x \in S$, in G/H has the limit point yH; i.e., G/H is compact. Conversely, we shall prove that the existence of a fundamental set with compact closure follows again from the compactness of G/H, provided G satisfies the second axiom of countability. However, in the following lemma, we do not need the latter assumption.

LEMMA 4: *Let the topological space G/H be compact; then G is locally compact.*

PROOF: Let U be a neighborhood of e in G containing no other element of H. Since the topological space G is regular, we can choose a neighborhood W of e such that $\bar{W}^{-1}WW^{-1}\bar{W} \subset U$. Let S be an infinite subset of $\bar{W}$. Since G/H is compact, there exists a point z in G such that, for any neighborhood V of z in G, the set VH contains infinitely many points of S. Let $V \subset Wz$, and let x, y be two points of $VH \cap S$; then xa and yb are points of Wz, for suitably selected elements a and b of H, and

$$ab^{-1} \in x^{-1}WW^{-1}y \subset \bar{W}^{-1}WW^{-1}\bar{W} \subset U;$$

hence $a = b$. It follows that there exists a fixed $a \in H$ such that Va contains infinitely many points of S, for any neighborhood V of z; then S has the limit point za, and $\bar{W}$ is compact. Hence G is locally compact; q.e.d.

Assume now that G satisfies the second axiom of countability, and consider again the construction of F in the proof of Lemma 2. In view of Lemma 4, we can suppose that the sets $\bar{V}_x$ are compact; then also the sets $\bar{B}_k$ and $\bar{F}_k$ are compact. We shall prove that F_k is empty for all sufficiently large values of k. Otherwise there would exist a sequence of points $x_1, x_2, \cdots$ in G such that $x_n \in F_{k_n}$, where $k_1, k_2, \cdots$ is an increasing sequence of indices, and a sequence of points $a_1, a_2, \cdots$ in H such that the sequence $x_n a_n (n = 1, 2, \cdots)$ converges to a point x_0 in G. The point x_0 lies in some B_k. If $x_0 \in B_l$ and $k > l$, then

$$B_l \cap F_k H = B_l \cap (B_k H - (B_1 \cup \cdots \cup B_{k-1})H) = 0;$$

consequently, no $x_n a_n (k_n > l)$ lies in B_l, and this is a contradiction. It follows that $F = F_1 \cup F_2 \cup \cdots$ is the union of a finite number of F_k; hence $\bar{F}$ is also compact.

Let $\bar{F}$ be compact, and let C be an arbitrary compact subset in G. Consider all $a \in H$ such that $\bar{F}a \cap C \neq 0$. All these a are contained in the compact set $\bar{F}^{-1}C$; since H is discrete, they belong to a finite set. On the other hand, the images Fa cover the whole space G. It follows that C is completely covered by a finite number of images Fa. Since G is locally compact, every point in G has a neighborhood contained in a finite number of images of F. Furthermore, for $C = \bar{F}$, we see that F has only a finite number of neighbors Fa.

9. In this section we assume that there exists a normal fundamental set F; this means that a suitably chosen neighborhood of every point is contained

in the union of a finite number of images of F. It is obvious that then every compact set in G is covered by a finite number of images of F. Let A be the set of all elements a of H which are contained in $\bar{F}^{-1}\bar{F}$; it is clear that $a \in A$, if and only if Fa is a neighbor of F. Let H_0 be the smallest group containing A; i.e., H_0 is the group generated by all $a \in A$. We are going to prove that $H_0 = H$, provided G is connected.

If $H_0 \neq H$, let $G_0 = \sum_{a \in H_0} Fa$; then $G_0 \neq 0$ and $G - G_0 = \sum_{b \in H - H_0} Fb \neq 0$. Since G is connected, the two sets $\bar{G}_0$ and $\overline{G - G_0}$ have a common point x. On the other hand, only a finite number of images of F enter into a suitably chosen neighborhood of x. It follows that there exist a point a in H_0 and a point b in $H - H_0$ such that $x \in \bar{F}a$ and $x \in \bar{F}b$; then $ba^{-1} \in \bar{F}^{-1}\bar{F}$, $ba^{-1} \in A \subset H_0$, $b \in H_0$, and this is a contradiction.

10. In this section we assume that G is locally compact and satisfies the second axiom of countability. The volume $v(B)$ of a Borel set B in G has the following properties:[8] $v(\sum_k B_k) = \sum_k v(B_k)$; $v(Bx) = v(B)$, for all $x \in G$; $v(xB) = \Delta(x)v(B)$, where $\Delta(x)$ is a positive continuous function of x and $\Delta(xy) = \Delta(x)\Delta(y)$, for all $x, y \in G$; $v(B) > 0$, for all open sets B; $v(B)$ is finite for all compact sets B. Moreover, by these properties, $v(B)$ is uniquely determined up to a positive constant factor.

Let E, F be two fundamental sets relative to H. Any Borel set in G/H can be expressed in the form BH, where B is a Borel set in G; let $BH \cap E = C$, $BH \cap F = D$. In view of Lemma 3,

$$C = \sum_{a \in H} (BH \cap E_a), \qquad D = \sum_{a \in H} (BH \cap F_a), \qquad F_a = E_a a$$

$$v(C) = \sum_{a \in H} v(BH \cap E_a), \qquad v(D) = \sum_{a \in H} v(BH \cap F_a), \qquad v(BH \cap E_a) = v(BH \cap F_a);$$

hence $v(C) = v(D)$, and we can define, for all Borel sets BH in G/H, a volume in the space G/H by the formula $v_H(B) = v(BH \cap F)$. Then $v_H(B) \leqq v(F)$, and $v_H(G) = v(F) = v(E)$.

LEMMA 5: *Let $v(F)$ be finite; then $v(B)$ and $v_H(B)$ are left-invariant.*

PROOF: It follows from the definition of a fundamental set F that also $E = xF$ is a fundamental set, for any given x in G. Hence

$$v(F) = v(E) = \Delta(x)v(F).$$

We infer from the construction of the particular fundamental set F in the proof of Lemma 2 that it contains the open set B_1; hence $v(F) > 0$. On the other hand, we assumed $v(F)$ to be finite. Therefore $\Delta(x) = 1$, for all $x \in G$, and

$$v_H(xB) = v(xBH \cap E) = \Delta(x)v(BH \cap F) = v_H(B).$$

---

[8] J. von Neumann, *The uniqueness of Haar's measure,* Rec. Math. N. S. (Mat. Sbornik) 1 (43), pp. 721–734 (1936). [English. Russian summary.]

11. LEMMA 6: *Let C be a compact subgroup of G; then the representation of H in $C\backslash G$ is discontinuous.*

PROOF: Let $a_1, a_2, \cdots$ be a sequence of distinct elements in H, let $x \in G$, and assume that the set of right cosets Cxa_n $(n = 1, 2, \cdots)$ has the limit point Cy in $C\backslash G$. We choose a neighborhood P of y and a neighborhood Q of e such that $P^{-1}Q^{-1}QP$ contains no element $\neq e$ of H. Then $xa_n = c_n p_n$, $c_n \in C$, and $p_n \in P$ for all sufficiently large n. Since C is compact, there exists a point $c \in C$ such that $c_n \in cQ$ for infinitely many n. Let $c_n \in cQ$, $c_m \in cQ$, $m \neq n$, then $a_m^{-1}a_n = p_m^{-1}c_m^{-1}c_n p_n \in P^{-1}Q^{-1}QP$; hence $a_m = a_n$, and this is a contradiction. It follows that the representation of H in $C\backslash G$ is discontinuous.

PROOF OF THEOREM 1: In view of Lemma 6 we have only to prove the necessity of the compactness of C. We assume that the representation of H in $C\backslash G$ is discontinuous, where C is a given closed subgroup of G. Consider the particular fundamental set F constructed in the proof of Lemma 2; it contains the open set B_1, hence it contains also an open set B_0, whose closure $\bar{B}_0 = B$ is compact. Then $v_H(B) = v(BH \cap F) \geqq v(B_0H \cap F) = v(B_0) > 0$. Consequently there exists a compact set B such that the volume $v_H(B)$ is positive.

Let $c_n (n = 1, 2, \cdots)$ be a sequence of distinct points in C, and define

$$c_n^{-1}BH \cap F = P_n, \qquad P_n \cup P_{n+1} \cup \cdots = Q_n \ (n = 1, 2, \cdots),$$
$$Q_1 \cap Q_2 \cap \cdots = Q.$$

Then P_n, Q_n, Q are Borel sets, $Q_n \subset F$, and by Lemma 5

$$v(F) \geqq v(Q_n) \geqq v(P_n) = v_H(c_n^{-1}B) = v_H(B);$$

moreover

$$\sum_{k=1}^{n-1} v(Q_k - Q_{k+1}) = v(Q_1) - v(Q_n) \leqq v(F) \qquad (n = 1, 2, \cdots)$$

$$\sum_{k=1}^{\infty} v(Q_k - Q_{k+1}) = v(Q_1) - v(Q);$$

hence

$$v(Q) = \lim_{n\to\infty} v(Q_n) \geqq v_H(B) > 0.$$

It follows that Q is non-empty.

Let $x \in Q$; then $x \in P_n$ for an infinite number of indices n, hence $x \in c_n^{-1}BH$ for the same set of indices. Consequently, there exists a sequence $a_n \in H (n = 1, 2, \cdots)$ such that $c_n x a_n \in B$ for infinitely many n. But B is compact and the representation of H in $C\backslash G$ is discontinuous. Therefore it follows from $Cxa_n \in CB$ that a_n belongs to a finite set of elements in H. This implies that $c_n xa \in B$, for a fixed $a \in H$ and infinitely many n, and hence the set $c_n(n = 1, 2, \cdots)$ has a limit point c_0. Since C is closed, c_0 is a point of C. This proves the compactness of C.

12. Let C be a compact subgroup of G. The representation of H in $C\backslash G$ has no fixed point, if and only if none of the conjugate subgroups $x^{-1}Cx$, $x \in G$, contains an element $a \neq e$ of H. This assumption is satisfied, in particular, if H does not contain any element of finite order except e: Let $a \in x^{-1}Cx$, $a \in H$, then $a^n \in x^{-1}Cx$ $(n = 1, 2, \cdots)$; since $x^{-1}Cx$ is compact and H is discrete, it follows that a is an element of finite order, hence $a = e$.

If a fundamental set F has the property $F = CF$, we may consider it also as a set in $C\backslash G$.

LEMMA 7: *Let G satisfy the second axiom of countability; then a fundamental set in $C\backslash G$ relative to H exists, if and only if the representation of H in $C\backslash G$ has no fixed point.*

PROOF: The necessity of the condition is easily proved: Let $F = CF$ be a fundamental set in $C\backslash G$ relative to H, and let $a \in x^{-1}Cx$, $a \in H$. Choose $b \in H$ such that $xb \in F$; then $b^{-1}ab \in (xb)^{-1}C(xb) \subset F^{-1}C\,F = F^{-1}F$, $F(b^{-1}ab) \cap F \neq 0$; hence $b^{-1}ab = e$, $a = e$. In this part of the proof, we did not use the assumption that G satisfies the second axiom of countability.

Conversely, assume that the representation of H in $C\backslash G$ has no fixed point. Let $x \in G$ and $y \in C$. Since $x^{-1}yx \neq a$, for all elements $a \neq e$ of H, there exist a neighborhood $P_{x,y}$ of x and a neighborhood $Q_{x,y}$ of y such that $P_{x,y}^{-1}Q_{x,y}P_{x,y}$ does not contain any element $\neq e$ of H. For any given x, the compact set C can be covered by a finite number of the neighborhoods $Q_{x,y}$, and the corresponding $P_{x,y}$ have a non-empty open intersection V_x. Consequently every point x in G has a neighborhood V_x such that $CV_xa \cap CV_x = 0$, for all $a \neq e$ in H.

Now we generalize the proof of Lemma 2 in the following way. We define

$$F_1 = CB_1, \qquad F_k = CB_k - (CB_1 \cup \cdots \cup CB_{k-1})H \qquad (k = 2, 3, \cdots),$$

$$F = F_1 \cup F_2 \cup \cdots .$$

Since $C(CA \cup CB) = CA \cup CB$ and $C(CA - CB) = CA - CB$, for arbitrary sets A and B, it is readily proved that F is a fundamental set and that $F = CF$.

13. In this section we assume that G is a locally compact group satisfying the second axiom of countability. Let C be a compact subgroup of G. Any Borel set in $C\backslash G$ can be expressed in the form CB, where B is a Borel set in G. The formula $v_c(B) = v(CB)$ defines in $C\backslash G$ a right-invariant completely additive volume, which is positive for open sets and finite for compact sets; it is uniquely determined by these properties, up to an arbitrary positive constant factor.[9] In order to establish that H is a subgroup of the first kind in G, it is sufficient to prove the existence of a normal fundamental set $F = CF$ in $C\backslash G$, of finite volume $v_c(F)$, having only a finite number of neighbors.

[9] A. Weil, *L'intégration dans les groupes topologiques et ses applications*, Paris (1940); pp. 42–45.

By the result of section 8, the group H is certainly of the first kind, whenever G/H is compact.

In the next chapter we shall apply the following lemma:

LEMMA 8: *The group H is of the first kind, if and only if there exists, for some compact subgroup C of G, an open set $M = CM$ in $C\backslash G$ of finite volume $v_c(M)$ such that $MH = G$ and that $Ma \cap M \neq 0$ holds only for a finite number of $a \in H$.*

PROOF: Let H be of the first kind; this means that there exists a normal fundamental set F of finite volume $v(F)$, having only a finite number of neighbors. Let Q be the union of all open sets contained in $\sum_{a \in A} \bar{F}a$, where A denotes the finite set of all elements of H contained in $\bar{F}^{-1}\bar{F}$; the set Q is open. If x is an arbitrary point of F, then a certain neighborhood W of x is covered by the union of a finite number of images Fa, $a \in H$; on the other hand, if Fa enters into every neighborhood of x then Fa is a neighbor of F and $a \in A$; consequently there exists also a neighborhood $W \subset Q$. It follows that $x \in Q$, $F \subset Q$, $QH = G$. Furthermore,

$$Qa \cap Q \subset \bigcup_{b,c \in A} (\bar{F}ba \cap \bar{F}c);$$

hence $Qa \cap Q = 0$ for any element $a \in H$ not lying in the finite set $A^{-1}AA$. Therefore $M = Q$ and $C = e$ have the required properties.

Conversely, let C be a compact subgroup of G, and let $M = CM$ be open, $v_c(M)$ finite, $MH = G$, $Ma \cap M \neq 0$ for only a finite set S of $a \in H$. In view of $MH = G$, we can determine, by the construction in the proof of Lemma 2, the fundamental set F as a subset of M; then $v(F) \leqq v(M) = v_c(M)$, hence $v(F)$ is finite. Let x be an arbitrary point in G, and let $x \in Mb$, $b \in H$; then b belongs to a finite set S_x. The intersection of all Mb, $b \in S_x$, is a neighborhood W_x of x. If $Fa \cap W_x \neq 0$, $a \in H$, then $Ma \cap W_x \neq 0$, $Ma \cap Mb \neq 0$, for $b \in S_x$; hence a belongs to the finite set SS_x. It follows that F is normal. Moreover, let $c \in H$, and let $\bar{M}c$, $\bar{M}$ have a common point x. We choose $a \in H$ such that $x \in Ma$; then $Ma \cap Mc \neq 0$ and $Ma \cap M \neq 0$, hence $ac^{-1} \in S$ and $a \in S$; this proves that c belongs to the finite set $S^{-1}S$. Since $\bar{F} \subset \bar{M}$, the fundamental set F has only a finite number of neighbors. Consequently the group H is of the first kind in G.

III. GROUPS OF UNITS IN SIMPLE ORDERS

14. Let A_0 be a simple algebra of finite rank in R_0, the field of rational numbers; then A_0 is isomorphic to an algebra of n-rowed matrices $\xi = (\xi_{kl})$, whose elements ξ_{kl} $(k, l = 1, \cdots, n)$ are arbitrary quantities in a division algebra D_0. The center Z of D_0 is an algebraic number field; let h denote the degree of Z over R_0, and let g be the rank of D_0 in R_0; then $g/h = s^2$ is the square of a positive rational integer s. We choose a basis $\delta_1, \cdots, \delta_g$ of D_0 in R_0 and extend R_0 into R, the field of real numbers; then D_0 is extended into a semi-simple algebra D in R. Every element of D is uniquely expressed by the linear form $\delta = \delta_1x_1 + \cdots + \delta_gx_g$ with arbitrary real $x_1, \cdots, x_g$; we say that $x_1, \cdots, x_g$ are

the coordinates of δ. In the regular representation of D, the element δ is represented by a g-rowed matrix $\hat{\delta}$. The g-rowed unit matrix will be denoted by ϵ.

Let r_2 be the number of imaginary infinite prime spots of Z, and let D_0 be ramified at r_3 real infinite prime spots of Z; define $r_1 = h - 2r_2 - r_3$. There exists a non-singular g-rowed matrix c with constant complex elements such that $c\hat{\delta}c^{-1}$ decomposes into s equal matrices λ of degree hs, and λ itself breaks up into h matrices $\lambda_1, \cdots, \lambda_h$ of degree s; the g elements of these matrices are homogeneous linear functions of $x_1, \cdots, x_g$, whose matrix is the inverse of c'; the r_1 matrices $\lambda_1, \cdots, \lambda_{r_1}$ are real; the r_2 pairs of matrices λ_{r_1+k} and $\lambda_{r_1+r_2+k}$ ($k = 1, \cdots, r_2$) are conjugate complex; the r_3 matrices λ_k ($k = r_1 + 2r_2 + 1, \cdots, h$) have the form $\begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix}$, where α and β denote complex matrices of degree $s/2$; moreover, the matrix $c'\bar{c}$ is real. In order to simplify the notation we write $c\hat{\delta}c^{-1} = \delta$.

The general quantity of the algebra A_0 in R, or A, is the n-rowed matrix $\xi = (\xi_{kl})$ with arbitrary elements ξ_{kl} from D; we define $(\hat{\xi}_{kl}) = \hat{\xi}$ and consider ξ as the n-rowed matrix with the g-rowed matrix elements $c\hat{\xi}_{kl}c^{-1} = \xi_{kl}$ ($k, l = 1, \cdots, n$). The gn^2 coordinates of the ξ_{kl} form the coordinates of ξ. Let G be the set of all ξ lying on the surface $|\xi| = \pm 1$. If we introduce the natural topology of R for all coordinates of ξ, the set G becomes a locally compact topological group satisfying the second axiom of countability. Let C be a *maximal* compact subgroup of G. Since the elements of C are matrices, there exists a positive hermitian matrix μ such that $\bar{\xi}'\mu\xi = \mu$ for all $\xi \in C$; it follows from the proof of existence of μ that μ can be chosen as an element of G. Then there exists $\eta \in G$ such that $\mu = \bar{\eta}'\eta$, and $\eta\xi\eta^{-1}$ is unitary. Let C_0 denote the group of all unitary elements of G; then C_0 is a compact subgroup of G, and any maximal compact subgroup of G is a conjugate subgroup $C = \eta^{-1}C_0\eta$, $\eta \in G$. Two points ξ_1, ξ_2 of G lie in the same right coset $C\xi_0$ of C, if and only if $\bar{\xi}_1'\mu\xi_1 = \bar{\xi}_2'\mu\xi_2$; consequently the topological space T of all positive hermitian matrices $\zeta = \bar{\xi}_0'\mu\xi_0$ in G is homeomorphic to the space $C\backslash G$. It is easily proved that the number of dimensions of T is $w_n - 1$, where $w_n = (hns + r_1 - r_3)ns/2$. In T the group G is represented by the transformations $\zeta \to \bar{\xi}'\zeta\xi$, $\xi \in G$. Let H be any discrete subgroup of the first kind in G; then the representation $\zeta \to \bar{\vartheta}'\zeta\vartheta$, $\vartheta \in H$, of H in T is discontinuous, and it follows from Theorem 1 that this representation is uniquely determined by the condition that the dimension of T be as small as possible.

15. Let an order J_1 in D_0 be given, and denote by J_n the order in A_0 consisting of all n-rowed matrices $\alpha = (\alpha_{kl})$, $\alpha_{kl} \in J_1$ ($k, l = 1, \cdots, n$). We assume now that the basis $\delta_1, \cdots, \delta_g$ of D_0 is a minimal basis of J_1; then an element of D lies in J_1, if and only if its coordinates are rational integers. We say that a column of n quantities $\xi_1, \cdots, \xi_n$ in D_0 is a *vector* $\mathfrak{x}$; if the *components* $\xi_1, \cdots, \xi_n$ lie in J_1, then $\mathfrak{x}$ is called *integral.*

Let $\zeta \in T$, and define $\bar{\mathfrak{x}}'\zeta\mathfrak{x} = \zeta[\mathfrak{x}]$, $\Phi(\mathfrak{x}) = \sigma(\zeta[\mathfrak{x}])$, where σ denotes the trace.

Obviously $\Phi(\mathfrak{x})$ is a positive quadratic form of gn variables, namely the coordinates of the components of the vector $\mathfrak{x}$. Let V be the volume of the ellipsoid $\Phi(\mathfrak{x}) < 1$, and determine $\eta \in G$ such that $\bar{\eta}'\eta = \zeta$; since the linear transformation $\eta\mathfrak{x} \to \mathfrak{x}$ has the determinant ± 1, it follows that V is independent of the point ζ in T. For our further purposes we do not need the exact value of V; a simple calculation, which we omit, leads to the formula

$$V = \pi^{\frac{gn}{2}} d^{-\frac{n}{2}} / \Gamma\left(1 + \frac{gn}{2}\right),$$

where d denotes the absolute value of the discriminant of the basis $\delta_1, \cdots, \delta_g$.

If $\mathfrak{x}_1, \cdots, \mathfrak{x}_k$ are vectors, then we denote by $L(\mathfrak{x}_1, \cdots, \mathfrak{x}_k)$ the set of all vectors $\mathfrak{x}_1 t_1 + \cdots + \mathfrak{x}_k t_k$ with arbitrary rational $t_1, \cdots, t_k$. We determine gn integral vectors $\mathfrak{y}_k (k = 1, \cdots, gn)$ by the condition that the minimum of $\Phi(\mathfrak{x})$, in the set of all integral $\mathfrak{x}$ outside $L(\mathfrak{y}_1, \cdots, \mathfrak{y}_{k-1})$, be attained for $\mathfrak{x} = \mathfrak{y}_k$; if $k = 1$, the set $L(\mathfrak{y}_1, \cdots, \mathfrak{y}_{k-1})$ is defined as the null-vector. Put $\Phi(\mathfrak{y}_k) = N_k$, then $N_1 \leqq N_2 \leqq \cdots$ and, by Minkowski's theorem,

$$\prod_{k=1}^{gn} N_k \leqq 4^{gn} V^{-2}.$$

16. If $\mathfrak{x}_1, \cdots, \mathfrak{x}_k$ are vectors, then we denote by $L^*(\mathfrak{x}_1, \cdots, \mathfrak{x}_k)$ the set of all vectors $\mathfrak{x}_1\tau_1 + \cdots + \mathfrak{x}_k\tau_k$ with arbitrary $\tau_1, \cdots, \tau_k$ in D_0. We determine n integral vectors $\mathfrak{x}_k$ $(k = 1, \cdots, n)$ by the condition that the minimum of $\Phi(\mathfrak{x})$, in the set of all integral $\mathfrak{x}$ outside $L^*(\mathfrak{x}_1, \cdots, \mathfrak{x}_{k-1})$, be attained for $\mathfrak{x} = \mathfrak{x}_k$; if $k = 1$, the set $L^*(\mathfrak{x}_1, \cdots, \mathfrak{x}_{k-1})$ is defined as the null-vector. We call $\mathfrak{x}_1, \cdots, \mathfrak{x}_n$ an *extremal set.* Put $\Phi(\mathfrak{x}_k) = M_k$, then $M_1 \leqq \cdots \leqq M_n$.

In order to find a relation between these $\mathfrak{x}_k$ and the $\mathfrak{y}_k$ of the preceding section, we use an idea of Weyl. Let $\mathfrak{z}_1, \cdots, \mathfrak{z}_q$ be any vectors in the set $L^*(\mathfrak{x}_1, \cdots, \mathfrak{x}_{k-1})$, and let $q > g(k - 1)$; then there exist q rational numbers $t_1, \cdots, t_q$, not all 0, such that $\mathfrak{z}_1 t_1 + \cdots + \mathfrak{z}_q t_q = 0$. Consequently, for any given positive index $k \leqq n$, there exists a uniquely determined positive index $l = l_k \leqq g(k - 1) + 1 \leqq gn$, such that $\mathfrak{y}_t$ lies in $L^*(\mathfrak{x}_1, \cdots, \mathfrak{x}_{k-1})$ for $t < l$ and outside for $t = l$. By the definition of $\mathfrak{x}_k$, we obtain the inequality $\Phi(\mathfrak{y}_l) \geqq \Phi(\mathfrak{x}_k)$. On the other hand, the vector $\mathfrak{x}_k$ lies outside $L^*(\mathfrak{x}_1, \cdots, \mathfrak{x}_{k-1})$ and, *a fortiori*, outside $L(\mathfrak{y}_1, \cdots, \mathfrak{y}_{l-1})$; hence $\Phi(\mathfrak{x}_k) \geqq \Phi(\mathfrak{y}_l)$. Consequently, $\Phi(\mathfrak{x}_k) = \Phi(\mathfrak{y}_l)$,

$$M_k^g \leqq \prod_{r=g(k-1)+1}^{gk} N_r, \qquad \prod_{k=1}^{n} M_k \leqq 4^n V^{-\frac{2}{g}}.$$

17. Let $\mathfrak{x}_1, \cdots, \mathfrak{x}_n$ be an extremal set, and let $(\mathfrak{x}_1 \cdots \mathfrak{x}_n) = \nu = \nu_\zeta$; we have to bear in mind that ν_ζ is not necessarily uniquely determined by ζ. It follows from the construction of $\mathfrak{x}_1, \cdots, \mathfrak{x}_n$ that $\mathfrak{x}_1\tau_1 + \cdots + \mathfrak{x}_n\tau_n \neq 0$, for any system of quantities $\tau_1, \cdots, \tau_n$ in D_0, not all 0; consequently $|\nu| \neq 0$, and $|\nu|^2 = N$ is a positive rational integer, $N \geqq 1$.

Put $\mathfrak{x} = \nu\mathfrak{y}$, then $\zeta[\mathfrak{x}] = \rho[\mathfrak{y}]$, where $\rho = \bar{\nu}'\zeta\nu = \zeta[\nu]$ is the n-rowed matrix

with the elements $\rho_{kl} = \bar{\mathfrak{x}}_k'\zeta\mathfrak{x}_l$ $(k, l = 1, \cdots, n)$ in D. Generalizing the Jacobi transformation, we have $\rho = \omega[\delta]$, where δ is a *triangular* matrix with the elements δ_{kl} $(k, l = 1, \cdots, n)$ in D, i.e., $\delta_{kl} = 0$ $(k > l)$ and $\delta_{kk} = \epsilon$, and ω is a diagonal matrix whose diagonal elements $\omega_1, \cdots, \omega_n$ lie in D. Let $\mathfrak{y} = (\eta_1, \cdots, \eta_n)'$, $\delta\mathfrak{y} = \mathfrak{z} = (\zeta_1 \cdots \zeta_n)'$, then

$$\zeta_k = \eta_k + \sum_{l=k+1}^{n} \delta_{kl}\eta_l \qquad (k = 1, \cdots, n).$$

Obviously the matrix $\omega_k (k = 1, \cdots, n)$ is positive hermitian. Since

$$\rho_{kk} = \omega_k + \sum_{l=1}^{k-1} \omega_l[\delta_{lk}] \qquad (k = 1, \cdots, n),$$

the hermitian matrix $\rho_{kk} - \omega_k$ is non-negative; hence

$$M_k = \Phi(\mathfrak{x}_k) = \sigma(\rho_{kk}) \geqq \sigma(\omega_k) > 0. \tag{1}$$

On the other hand,

$$N = |\nu|^2 = |\zeta[\nu]| = |\rho| = |\omega| = \prod_{k=1}^{n} |\omega_k|,$$

and consequently

$$N^{\frac{1}{g}} \prod_{k=1}^{n} \frac{\sigma(\omega_k)}{|\omega_k|^{\frac{1}{g}}} \prod_{k=1}^{n} \frac{M_k}{\sigma(\omega_k)} = \prod_{k=1}^{n} M_k \leqq 4^n V^{-\frac{2}{g}}. \tag{2}$$

Let $r_1, \cdots, r_g$ be the characteristic roots of the matrix ω_k; then

$$\sigma(\omega_k) = r_1 + \cdots + r_g \geqq g(r_1 \cdots r_g)^{\frac{1}{g}} = g\,|\omega_k|^{\frac{1}{g}}. \tag{3}$$

It follows from (1), (2), (3) that the numbers N, $\sigma(\omega_k)/|\omega_k|^{\frac{1}{g}}$, $M_k/\sigma(\omega_k)$ are bounded in T; then also the quotients r_p/r_q $(p, q = 1, \cdots, g)$ and, by the inequality $M_k \leqq M_{k+1}$, the quotients $\sigma(\omega_k)/\sigma(\omega_{k+1})$ $(k = 1, \cdots, n-1)$ are bounded in T.

Let k and l be given indices, $1 \leqq k < l \leqq n$, and choose $\eta_l = \epsilon$, $\eta_p = 0 (k < p \neq l)$, hence $\zeta_p = \delta_{pl}$ $(k < p < l)$, $\zeta_l = \epsilon$, $\zeta_p = 0$ $(p > l)$; furthermore, determine η_p $(p \leqq k)$ in J_1 by the condition that the coordinates of

$$\zeta_p = \eta_p + \sum_{q=p+1}^{n} \delta_{pq}\eta_q \qquad (p = k, k-1, \cdots, 1)$$

lie in the real interval $-\frac{1}{2} \leqq x < \frac{1}{2}$. Since

$$\mathfrak{x} = \nu\mathfrak{y} = \mathfrak{x}_1\eta_1 + \cdots + \mathfrak{x}_k\eta_k + \mathfrak{x}_l$$

is an integral vector outside $L^*(\mathfrak{x}_1, \cdots, \mathfrak{x}_{l-1})$, it follows that $\Phi(\mathfrak{x}) \geqq \Phi(\mathfrak{x}_l)$,

$$\sum_{p=1}^{k} \sigma(\omega_p[\zeta_p]) \geqq \sum_{p=1}^{k} \sigma(\omega_p[\delta_{pl}]);$$

but ζ_p $(p = 1, \cdots, k)$ is bounded, and consequently the quotients $\sigma(\omega_k[\delta_{kl}])/\sigma(\omega_k)$ $(1 \leqq k < l \leqq n)$ are bounded. Moreover there exists a unitary matrix u such that $\omega_k = \bar{u}'ru$, where r is the g-rowed diagonal matrix with the diagonal elements $r_1, \cdots, r_g$. It follows that δ_{kl} $(1 \leqq k < l \leqq n)$ is bounded in T.

18. Let $\mathfrak{P}$ be an m-rowed positive symmetric matrix with real elements. By the Jacobi transformation we have a uniquely determined representation $\mathfrak{P} = \mathfrak{Q}[\mathfrak{D}] = \mathfrak{D}'\mathfrak{Q}\mathfrak{D}$, where $\mathfrak{D} = (d_{kl})$ $(k, l = 1, \cdots, m)$ is a real triangular matrix, i.e., $d_{kl} = 0$ $(k > l)$ and $d_{kk} = 1$, and $\mathfrak{Q}$ is a diagonal matrix with positive diagonal elements $q_1, \cdots, q_m$. For any positive t, the $\frac{m(m+1)}{2} - 1$ inequalities

$$q_k/q_{k+1} < t, \qquad -t < d_{kl} < t \qquad\qquad (1 \leqq k < l \leqq m)$$

define an open set Π_t in the space of all $\mathfrak{P}$.

Consider now the positive hermitian matrix

$$\mathfrak{P}_\rho = (\bar{c}'\rho_{kl}c) = \omega[(\delta_{kl}c)] = \omega[(c\hat{\delta}_{kl})];$$

since $\bar{c}'\omega_k c = \bar{c}'c\hat{\omega}_k$ is real, the matrices $\bar{c}'\omega_k c$ and $\mathfrak{P}_\rho$ are real positive symmetric· It follows from the results of the preceding section that $\mathfrak{P}_\rho \in \Pi_f$, where the positive number f does not depend upon the point $\zeta = \rho[\nu^{-1}]$ in T.

Let H_0 be the group of all units ϑ in the order J_n; obviously H_0 is a discrete subgroup of G. The matrix $\nu = \nu_\zeta$ depends upon the point ζ in T. Since ν lies in J_n and $|\nu|$ is a bounded rational integer $\neq 0$, it follows that all ν lie in a finite number of right cosets of H_0 relative to the group xG, where x is any positive real number. Let $S = \nu_1, \cdots, \nu_p$ be a complete set of representatives of these cosets; then $\nu = \vartheta\nu_k$, $\nu_k \in S$, $\vartheta \in H_0$.

The point ζ of T is called *reduced*, whenever $\nu_\zeta \in S$; let Q be the set of all reduced ζ. The *images* $Q[\vartheta] = \bar{\vartheta}'Q\vartheta$, $\vartheta \in H_0$, cover the whole space T.

Let ζ be a reduced point, and let $\nu_\zeta = \nu \in S$. Since Π_f is open, there exists a neighborhood W_ζ of ζ in T such that $\mathfrak{P}_\rho \in \Pi_f$, for $\rho = \zeta_0[\nu]$ and all $\zeta_0 \in W_\zeta$. Let W_0 be the union of all these W_ζ, $\zeta \in Q$. Then W_0 is open, $Q \subset W_0$; moreover, for any point ζ_0 of W_0, there exists an element $\nu \in S$ such that $\mathfrak{P}_\rho \in \Pi_f$, for $\rho = \zeta_0[\nu]$. Assume now that also $\zeta_0[\vartheta^{-1}] = \zeta^* \in W_0$, for some $\vartheta \in H_0$, and determine $\nu^* \in S$ such that $\mathfrak{P}_{\rho^*} \in \Pi_f$, for $\rho^* = \zeta^*[\nu^*]$. Choose a positive rational integer a such that $a\nu_k^{-1} \in J_n$ $(k = 1, \cdots, p)$ and define $a\nu^{-1}\vartheta^{-1}\nu^* = \beta$, then $\beta \in J_n$, $|\beta| = \pm |a\nu^{-1}\nu^*|$, $\mathfrak{P}_\rho[\hat{\beta}] = a^2\mathfrak{P}_{\rho^*}$. But also $a^2\mathfrak{P}_{\rho^*}$ lies in Π_f, and $\hat{\beta}$ is a gn-rowed matrix with integral rational elements and bounded determinant. By a known theorem concerning the reduction of positive quadratic forms, it follows that β belongs to a finite set, independent of ζ; hence also ϑ belongs to a finite set. Consequently we see that $W_0[\vartheta] \cap W_0 \neq 0$ only for a finite number of $\vartheta \in H_0$.

19. In this section we shall prove that Q is contained in an open set of finite volume.

Let P be any open set in T, and let z be a positive number. We denote by $P(z)$ the set of all products $p = x\zeta$, where $0 < x < z$ and $\zeta \in P$. Let $v_z(P)$ be the euclidean volume of $P(z)$, computed in terms of the coordinates of the points $p \in P(z)$; obviously $v_z(P) = v_1(P)z^{a_n}$ with $a_n = w_n/gn$, where w_n is the number of dimensions of $P(z)$. For any $\xi \in G$, the linear transformation $p \rightarrow p[\xi]$ of the coordinates has the determinant 1; consequently $v_1(P)$ defines the invariant volume in T, up to a positive constant factor.

By the results of section 17, it is sufficient to prove the finiteness of $v_1(P_n)$, where P_n consists of all $\rho = \omega[\delta]$ in T satisfying

$$(4)\quad \sigma(\omega_k) < t \mid \omega_k \mid^{\frac{1}{\sigma}} \; (k = 1, \cdots, n), \qquad \sigma(\omega_k)/\sigma(\omega_{k+1}) < t \qquad (k = 1, \cdots, n-1),$$
$$\sigma(\bar{\delta}'_{kl}\delta_{kl}) < t \qquad (1 \leqq k < l \leqq n),$$

and t denotes an arbitrarily given positive number. The points of $P_n(1)$ are defined by (4) and the condition $\mid \rho \mid < 1$. Obviously the coordinates of $\omega_1 = \rho_{11}$ are bounded. We apply induction with respect to n. The assertion is trivial for $n = 1$, since then all coordinates of ρ are bounded. Let $n \geqq 2$, and let the assertion be proved for $n - 1$ instead of n.

We have

$$\rho[\mathfrak{y}] - \omega_1[\zeta_1] = \sum_{k=2}^{n} \omega_k[\zeta_k],$$

with $(\zeta_1 \cdots \zeta_n)' = \delta\mathfrak{y}$. For any fixed values of the coordinates of $\rho_{1k} = \omega_1\delta_{1k}$ $(k = 1, \cdots, n)$, let V_1 be the euclidean volume of the corresponding w_{n-1}-dimensional surface of section in $P_n(1)$. Since

$$\prod_{k=2}^{n} \mid \omega_k \mid = \mid \omega_1 \mid^{-1} \mid \rho \mid < \mid \omega_1 \mid^{-1},$$

we have

$$V_1 < v_1(P_{n-1}) \mid \omega_1 \mid^{-a_{n-1}}.$$

Instead of the coordinates of ρ_{1k} $(k = 1, \cdots, n)$ we introduce the coordinates of ω_1 and $\delta_{1k} = \omega_1^{-1}\rho_{1k}$ $(k = 2, \cdots, n)$, as variables of integration; these coordinates are bounded. Since the functional determinant of the transformation is $\mid \omega_1 \mid^{n-1}$, the integrand becomes

$$V_1 \mid \omega_1 \mid^{n-1} < v_1(P_{n-1}) \mid \omega_1 \mid^{n-a_{n-1}-1};$$

but $v_1(P_{n-1})$ is finite and

$$n - a_{n-1} = n - \frac{1}{2hs}\,(h(n-1)s + r_1 - r_3) =$$
$$\frac{n+1}{2} - \frac{r_1 - r_3}{2hs} \geqq \frac{n+1}{2} - \frac{1}{2s} \geqq \frac{n}{2} \geqq 1;$$

hence the integrand is bounded, and $v_1(P_n)$ is finite.

20. By the results of the two preceding sections, there exists in T an open set W of finite volume, such that the images $W[\vartheta]$, $\vartheta \in H_0$, cover T and that $W[\vartheta] \cap W \neq 0$ holds only for a finite number of $\vartheta \in H_0$. Let M be the set of all $\xi \in G$ such that $\mu[\xi] = \zeta \in W$; then M fulfills the conditions of Lemma 8. It follows that H_0 is a group of the first kind in G.

Let J be an arbitrary order in A_0, and let H be the group of units in J. Then $J^* = J \cap J_n$ is again an order, and the group H^* of units in J^* is a subgroup of finite indices j and j_0 in H and H_0. Let

$$H_0 = \sum_{k=1}^{j_0} a_k H^*, \qquad H = \sum_{l=1}^{j} H^* b_l, \qquad M_0 = \bigcup_k M a_k;$$

then

$$M_0 H = \bigcup_{k,l} M a_k H^* b_l = \bigcup_l M H_0 b_l = M H_0 = G,$$

and $v(M_0) \leqq j_0 v(M)$ is finite. If ξ is a common point of M_0 and $M_0\vartheta$, $\vartheta \in H$, then $\mu[\xi] = \zeta$ is a common point of $W[a_k]$ and $W[a_l\vartheta]$, for some $k, l \leqq j_0$. Exactly as in section 18, it can be proved that ϑ belongs to a finite set independent of ξ, if W is suitably chosen. Consequently M_0 satisfies the conditions of Lemma 8. It follows that H is a group of the first kind in G, and Theorem 2 is proved.

The explicit construction of a normal fundamental set in G relative to H_0, and more generally relative to H, is perhaps of minor interest; therefore we give only the following sketch: Let ζ_0 be a frontier point of Q; then it is easily seen, from the definition of Q, that there exist for ζ_0 two extremal sets $\mathfrak{x}_1, \cdots, \mathfrak{x}_n$ and $\mathfrak{y}_1, \cdots, \mathfrak{y}_n$, and an index $l \leqq n$, such that the differences $L_k(\zeta) = \Phi(\mathfrak{x}_k) - \Phi(\mathfrak{y}_k)$ vanish identically in ζ for $k = 1, \cdots, l-1$, but not for $k = l$, and that $L_l(\zeta_0) = 0$; moreover, $\mathfrak{x}_l$ and $\mathfrak{y}_l$ belong to a finite set independent of ζ_0; hence the frontier of Q lies on a finite number of planes. Let S be the set of all $\xi \in G$ such that $\mu[\xi] = \zeta$ is reduced; then the frontier of S lies on a finite number of surfaces of the second order; furthermore, if $S\vartheta \cap S \neq 0$, $\vartheta \in H_0$, then ϑ belongs to a finite set. Let F_0 be the set of all $\xi \in S$ such that $\sigma(\xi\vartheta) \geqq \sigma(\xi)$, whenever $\xi\vartheta \in S$ and $\vartheta \in H_0$; then the frontier of F_0 lies on a finite number of surfaces of the second order, and F_0 is a fundamental domain relative to H_0; i.e., the images $F_0\vartheta$, $\vartheta \in H_0$, cover G completely without overlappings, common frontier points excepted. Omitting a suitably selected set of frontier points of F_0, we obtain a normal fundamental set F relative to H_0, having only a finite number of neighbors. Finally, the passage from F to a fundamental set relative to H may easily be performed.

The group G is not connected; it consists of the two sets G_1 and G_2 defined by $|\xi| = 1$ and $|\xi| = -1$, each of them being connected. The group $G_1 \cap H$ is either H itself or an invariant subgroup of index 2. It follows from the result of section 9 that H has a finite system of generators.

THE INSTITUTE FOR ADVANCED STUDY

ANNALS OF MATHEMATICS
Vol. 44, No. 4, October, 1943

ON THE THEORY OF LOCAL RINGS

BY CLAUDE CHEVALLEY

(Received August 12, 1943)

The present paper contains a certain number of purely algebraic results to be used in a local theory of algebraic varieties, which will be developed in a later paper.

The most important part of this paper is contained in §IV, where is defined a notion of "multiplicity" (which might also have been called degree of ramification) in a complete local ring. This notion of multiplicity will appear eventually as the main tool in the definition of the multiplicity of intersection of two algebraic or algebroid varieties.

Conventions of terminology

By a ring we mean always a comutative ring with unit element. When a ring $\mathfrak{o}$ is said to be a sub-ring of a ring $\mathfrak{o}'$, it is understood tacitly that the unit element of $\mathfrak{o}$ is also the unit element of $\mathfrak{o}'$ (except when the contrary is explicitly specified). If $\mathfrak{o}$ is a sub-ring of $\mathfrak{o}'$ and if $\mathfrak{a}$ is an ideal in $\mathfrak{o}$, we denote by $\mathfrak{a}\mathfrak{o}'$ the ideal generated in $\mathfrak{o}'$ by the elements of $\mathfrak{a}$.

A ring $\mathfrak{o}$ is said to be *Noetherian* if the maximal condition holds for the ideals in $\mathfrak{o}$, i.e. if every ideal has a finite set of generators.

If $\mathfrak{a}$ is an ideal in a Noetherian ring $\mathfrak{o}$, we call "prime divisors" of $\mathfrak{a}$ the prime ideals which occur as associated prime ideals of the primary ideals in a shortest representation of $\mathfrak{a}$ as intersection of primary ideals.

We call "*adherence*" of a set E in a topological space the set which is usually called the closure of E. Any point of this set is said to be adherent to E.

The symbol $\varnothing$ denotes the empty set.

§I. Rings of Quotients[0]

It is well known that any ring $\mathfrak{o}$ can be imbedded in a "ring of quotients" whose elements are the fractions of the form a/b, $a \in \mathfrak{o}$, b not being a zero-divisor in $\mathfrak{o}$.

Let now S be any set of non-zero divisors in $\mathfrak{o}$ which is multiplicatively closed (i.e. the product of two elements of S is in S). Those elements of the ring of quotients which may be written in the form a/b, with $a \in \mathfrak{o}$, $b \in S$ form a sub-ring $\mathfrak{o}_S$ of the field of quotients, which we shall call the *ring of quotients* of the set S.

LEMMA 1. *If $\mathfrak{a}$ is an ideal in $\mathfrak{o}_S$, we have $\mathfrak{a} = (\mathfrak{a} \cap \mathfrak{o})\cdot\mathfrak{o}_S$.*

The inclusion $(\mathfrak{a} \cap \mathfrak{o})\mathfrak{o}_S \subset \mathfrak{a}$ is obvious. Conversely, let a/b be any element of $\mathfrak{a}$, with $a \in \mathfrak{o}$, $b \in S$; we have $a = b.\ (a/b) \in \mathfrak{a} \cap \mathfrak{o}$, whence $a/b \in (\mathfrak{a} \cap \mathfrak{o})\mathfrak{o}_S$, which proves Lemma 1.

---

[0] The notion of ring of quotients of a set S is due to H. Grell (cf. *Beziehungen zwischen d. Idealen verschiedener Ringe*, Math. Ann. 97, 1926). The results contained in §I are not new (cf. Krull, *Idealtheorie*, Erg. d. Math.).

It follows immediately from Lemma 1 that $\mathfrak{o}_S$ is Noetherian when $\mathfrak{o}$ is itself Noetherian.

LEMMA 2. *Let $\mathfrak{p}$ be a prime ideal in $\mathfrak{o}$, and let $\mathfrak{q}$ be a primary ideal for $\mathfrak{p}$. If $\mathfrak{p} \cap S \neq \emptyset$, we have $\mathfrak{p}\mathfrak{o}_S = \mathfrak{q}\mathfrak{o}_S = \mathfrak{o}_S$. If $\mathfrak{p} \cap S = \emptyset$, the ideal $\mathfrak{p}\mathfrak{o}_S$ is prime in $\mathfrak{o}_S$; the ideal $\mathfrak{q}\mathfrak{o}_S$ is primary for $\mathfrak{p}\mathfrak{o}_S$; we have $\mathfrak{q}\mathfrak{o}_S \cap \mathfrak{o} = \mathfrak{q}$, and the primary ideals $\mathfrak{q}\mathfrak{o}_S$, $\mathfrak{q}$ have the same length.*

If $\mathfrak{p} \cap S \neq \emptyset$, we have also $\mathfrak{q} \cap S \neq 0$ (because any power of an element of S is in S), whence $\mathfrak{p}\mathfrak{o}_S = \mathfrak{q}\mathfrak{o}_S = \mathfrak{o}_S$. Assume now that $\mathfrak{p} \cap S = \emptyset$; then 1 cannot belong to $\mathfrak{p}\mathfrak{o}_S$, whence $\mathfrak{p}\mathfrak{o}_S \neq \mathfrak{o}_S$[1] and, *a fortiori*, $\mathfrak{q}\mathfrak{o}_S \neq \mathfrak{o}_S$. Let a/b, a'/b' be two elements of $\mathfrak{o}_S$ such that $a/b \cdot a'/b' \in \mathfrak{q}\mathfrak{o}_S$, $a'/b' \notin \mathfrak{q}\mathfrak{o}_S$ (with $a, a' \in \mathfrak{o}$, $b', b \in S$, $a' \notin \mathfrak{q}$). We have $aa'/bb' = c/c'$, $c \in \mathfrak{q}$, $c' \in S$, whence $c'aa' = cbb' \in \mathfrak{q}$; since $\mathfrak{p} \cap S = \emptyset$, c' does not belong to $\mathfrak{p}$, whence $aa' \in \mathfrak{q}$, and therefore $a^r \in \mathfrak{q}$ for some r; it follows that $(a/b)^r \in \mathfrak{q}\mathfrak{o}_S$, which proves that $\mathfrak{q}\mathfrak{o}_S$ is primary. If $\mathfrak{q} = \mathfrak{p}$, we can furthermore take $r = 1$, which proves that $\mathfrak{p}\mathfrak{o}_S$ is prime.

We have $\mathfrak{q}\mathfrak{o}_S \subset \mathfrak{p}\mathfrak{o}_S$; conversely, if $a/b \in \mathfrak{p}\mathfrak{o}_S$ ($a \in \mathfrak{p}$, $b \in S$), we have $a^r \in \mathfrak{q}$ for some r, whence $(a/b)^r \in \mathfrak{q}\mathfrak{o}_S$, which proves that $\mathfrak{q}\mathfrak{o}_S$ is primary for $\mathfrak{p}\mathfrak{o}_S$.

Let a be an element of $\mathfrak{q}\mathfrak{o}_S \cap \mathfrak{o}$; we may write a in the form b/c, $b \in \mathfrak{q}$, $c \in S$. Since $c \notin \mathfrak{p}$, the relation $ac \in \mathfrak{q}$ implies $a \in \mathfrak{q}$, whence $\mathfrak{q}\mathfrak{o}_S \cap \mathfrak{o} = \mathfrak{q}$. It follows that there exists a one-to-one inclusion preserving correspondence between the primary ideals for $\mathfrak{p}$ in $\mathfrak{o}$ and the primary ideals for $\mathfrak{p}\mathfrak{o}_S$ in $\mathfrak{o}_S$, which proves that $\mathfrak{q}$ and $\mathfrak{q}\mathfrak{o}_S$ have the same length.

LEMMA 3. *Assume that $\mathfrak{o}$ is Noetherian; let $\mathfrak{a} = (\bigcap_i \mathfrak{q}_i) \cap (\bigcap_j \mathfrak{d}_j)$ be an irredundant*[2] *representation of $\mathfrak{a}$ as an intersection of primary ideals, where we have separated the primary ideals $\mathfrak{q}_i$ with prime associated ideals $\mathfrak{p}_i$ such that $\mathfrak{p}_i \cap S = \emptyset$ from those $\mathfrak{d}_j$ whose associated prime ideals meet the set S. Then $\mathfrak{a}\mathfrak{o}_S = \bigcap_i \mathfrak{q}_i\mathfrak{o}_S$, and this is an irredundant representation of $\mathfrak{a}\mathfrak{o}_S$ as intersection of primary ideals in $\mathfrak{o}_S$.*

We have obviously $\mathfrak{a}\mathfrak{o}_S \subset \bigcap_i \mathfrak{q}_i\mathfrak{o}_S$; let conversely a be an element of $\bigcap_i \mathfrak{q}_i\mathfrak{o}_S$; for each i we may write $a = b_i/c_i$, $b_i \in \mathfrak{q}_i$, $c_i \in S$. For each index j corresponding to a $\mathfrak{d}_j$, we can find an element c_j in $\mathfrak{d}_j \cap S$; we set $c = \prod_i c_i \cdot \prod_j c_j$; we have $c \in S$ and $a = b/c$, with $b \in (\bigcap_i \mathfrak{q}_i) \cap (\bigcap_j \mathfrak{d}_j)$ whence $b \in \mathfrak{a}$ and $a \in \mathfrak{a}\mathfrak{o}_S$. In order to prove that the representation $\mathfrak{a}\mathfrak{o}_S = \bigcap_i \mathfrak{q}_i\mathfrak{o}_S$ is irredundant, we observe first that it follows from Lemma 2 that the prime divisors associated with the primary ideals $\mathfrak{q}_i$ are distinct from each other; moreover an inclusion $\mathfrak{q}_i\mathfrak{o}_S \subset \bigcap_{i' \neq i} \mathfrak{q}_{i'}\mathfrak{o}_S$ would imply by Lemma 1 that $\mathfrak{q}_i \subset \bigcap_{i' \neq i} \mathfrak{q}_{i'}$: this is impossible.

Let us now consider the case where $\mathfrak{o}$ does not contain any zero-divisor $\neq 0$ and where S is the set of the elements of $\mathfrak{o}$ which do not belong to some given prime ideal $\mathfrak{p}$. This set is clearly multiplicatively closed; the corresponding ring $\mathfrak{o}_S$ takes then the name of "quotient ring" of the ideal $\mathfrak{p}$ with respect to the ring $\mathfrak{o}$; we shall denote this ring by $\mathfrak{o}_\mathfrak{p}$. It is clear that the factor ring $\mathfrak{o}_\mathfrak{p}/\mathfrak{p}\mathfrak{o}_\mathfrak{p}$ is the field of quotients of $\mathfrak{o}/\mathfrak{p}$.

---

[1] Observe that, if $\mathfrak{a}$ is any ideal in $\mathfrak{r}$, the elements of $\mathfrak{a}\mathfrak{o}_s$ are of the form ax, $a \in \mathfrak{a}$, $x \in \mathfrak{r}$.

[2] This means that none of the ideals $\mathfrak{a}_i$, $\mathfrak{a}_j$ is contained in the intersection of the others.

Let us return to the general case, where S is any multiplicatively closed set of regular elements of a ring $\mathfrak{o}$. If $\mathfrak{p}$ is a prime ideal in $\mathfrak{o}$ which does not meet S, it follows from Lemma 2 that $\mathfrak{o}/\mathfrak{p}$ may be identified with a sub-ring of $\mathfrak{o}_S/\mathfrak{p}\mathfrak{o}_S$; it is clear that $\mathfrak{o}_S/\mathfrak{p}\mathfrak{o}_S$ has the same field of quotients as $\mathfrak{o}/\mathfrak{p}$; but, in general, $\mathfrak{o}_S/\mathfrak{p}\mathfrak{o}_S$ will not be itself a field.

§II. Semi-local Rings

DEFINITION 1. *A Noetherian ring $\mathfrak{o}$ is called a semi-local ring if there exist only a finite number of maximal prime ideals in $\mathfrak{o}$.*

Let $\mathfrak{p}_1, \cdots, \mathfrak{p}_h$ be all the distinct maximal prime ideals in a semi-local ring $\mathfrak{o}$. We set $\mathfrak{a} = \mathfrak{p}_1 \cdots \mathfrak{p}_h$. We shall define a topology in the ring $\mathfrak{o}$; a subset U of $\mathfrak{o}$ will be called an open set in this topology if it satisfies the following condition: if $u \in U$, there exists an exponent n (depending upon u) such that $u + \mathfrak{a}^n \subset U$. It is obvious that any union or finite intersection of open sets in the sense of this definition are open; the empty set and $\mathfrak{o}$ itself are open. The usual requirements for the definition of a topology are therefore satisfied. If u is any element of $\mathfrak{o}$, the sets $u + \mathfrak{a}^n$ form a fundamental system of neighbourhoods of u; if we observe that $(u + \mathfrak{a}^n) - (v + \mathfrak{a}^n) = u - v + \mathfrak{a}^n$, $(u + \mathfrak{a}^n)(v + \mathfrak{a}^n) \subset uv + \mathfrak{a}^n$, we see that the operations of addition, subtraction, multiplication in $\mathfrak{o}$ are continuous. Whenever we shall apply topological notions to a semi-local ring, it will be tacitly understood that these notions apply to the topology we have just defined.

We shall now prove that the Hausdorf separation axiom holds in our topology. It will be sufficient to prove that $\bigcap_{n=1}^{\infty}\mathfrak{a}^n = \{0\}$.

LEMMA 1. *If $\mathfrak{o}$ is a semi-local ring, an element of $\mathfrak{o}$ which does not belong to any one of the maximal prime ideals in $\mathfrak{o}$ is a unit in $\mathfrak{o}$.*

In fact, if we had $\mathfrak{o}u \neq \mathfrak{o}$, $\mathfrak{o}u$ would be contained in some prime ideal of $\mathfrak{o}$, and therefore also in some maximal prime ideal, which is not the case.

Taking Lemma 1 into account, the formula $\bigcap_{n=1}^{\infty}\mathfrak{a}^n = \{0\}$ follows immediately from

LEMMA 2. *Let $\mathfrak{a}$ be an ideal in a Noetherian ring $\mathfrak{o}$. The following statements are equivalent:* 1) *no element which is* $\equiv 1$ (mod. $\mathfrak{a}$) *is a zero divisor in* $\mathfrak{o}$; 2) *we have* $\bigcap_{n=1}^{\infty}\mathfrak{a}^n = \{0\}$.[3]

Let $\mathfrak{n}$ be the ideal $\bigcap_{n=1}^{\infty}\mathfrak{a}^n$. We shall prove that $\mathfrak{n} = \mathfrak{n}\mathfrak{a}$. Let $\mathfrak{n}\mathfrak{a} = \bigcap_i \mathfrak{q}_i$ be a representation of $\mathfrak{n}\mathfrak{a}$ as an intersection of primary ideals, and let $\mathfrak{p}_i$ be the prime divisor of $\mathfrak{q}_i$. If $\mathfrak{a} \not\subset \mathfrak{p}_i$, we have $\mathfrak{q}_i : \mathfrak{a} = \mathfrak{q}_i$; in fact, let a be an element of $\mathfrak{a}$ not contained in $\mathfrak{p}_i$, and let q be an element of $\mathfrak{q}_i : \mathfrak{a}$; we have $aq \in \mathfrak{q}_i$, $a \notin \mathfrak{p}_i$, whence $q \notin \mathfrak{q}_i$. Since $\mathfrak{n}\mathfrak{a} \subset \mathfrak{q}_i$, we have $\mathfrak{n} \subset \mathfrak{q}_i$ if $\mathfrak{a} \not\subset \mathfrak{p}_i$. This last conclusion holds also if $\mathfrak{a} \subset \mathfrak{p}_i$; in fact, we have $\mathfrak{p}_i^k \subset \mathfrak{q}_i$ for some k, whence $\mathfrak{n} \subset \mathfrak{a}^k \subset \mathfrak{p}_i^k \subset \mathfrak{q}_i$. It follows that $\mathfrak{n} \subset \bigcap_i \mathfrak{q}_i = \mathfrak{n}\mathfrak{a}$, whence $\mathfrak{n} = \mathfrak{n}\mathfrak{a}$. Let us now take a set of generators $\nu_1, \cdots, \nu_r$ of $\mathfrak{n}$; we have $\nu_i = \sum_{j=1}^{r} a_{ij}\nu_j$ with $a_{ij} \in \mathfrak{a}$; the determinant δ of the matrix $(\delta_{ij} - a_{ij})$ is $\equiv 1$ (mod. $\mathfrak{a}$) and we have $\delta\nu_i = 0$ $(1 \leqq i \leqq r)$; there-

[3] Cf. *Krull*, Dimensionstheorie in Stellenringen, J.f.d.r.u.M. 179, 1938, pp. 204–226.

fore, if 1) holds, we have $\nu_i = 0$ $(1 \leqq i \leqq r)$, whence $\mathfrak{n} = \{0\}$. Conversely, let us assume that there exist two elements a, b such that $a \in \mathfrak{a}$, $b \neq 0$, $(1 + a)b = 0$; we have $b = -ab = (-1)^n a^n b \in \mathfrak{a}^n$ for every n, whence $b \in \mathfrak{n}$, $\mathfrak{n} \neq \{0\}$.

Since the topology of $\mathfrak{o}$ is a Hausdorf topology, we may use the notion of limit. Since the first countability axiom obviously holds in $\mathfrak{o}$, it will be sufficient to consider limits of sequences. We shall say that a sequence (u_n) is "convergent" if $\lim_{n\to\infty}(u_{n+1} - u_n) = 0$, i.e. if, given any $k > 0$, we have $u_{n+1} - u_n \in \mathfrak{a}^k$ as soon as n is sufficiently large.

DEFINITION 2. *The semi-local ring* $\mathfrak{o}$ *is said to be complete if every convergent sequence of elements of* $\mathfrak{o}$ *has a limit in* $\mathfrak{o}$.

LEMMA 3. *Let* $\mathfrak{p}_1, \cdots, \mathfrak{p}_h$ *be the maximal prime ideals in a semi-local ring* $\mathfrak{o}$. *Let* $a_1, \cdots, a_h$ *be h given elements of* $\mathfrak{o}$, *and let n be a given exponent. The system of congruences* $x \equiv a_i \pmod{\mathfrak{p}_i^n}$ *has always a solution x, which is uniquely determined modulo* $(\mathfrak{p}_1 \cdots \mathfrak{p}_h)^n$.

It is clear that $\mathfrak{p}_i \not\subset \mathfrak{p}_j$ for $j \neq i$. Let $\mathfrak{a}_i$ be the ideal $\prod_{j\neq i}\mathfrak{p}_j$; we have $\mathfrak{a}_i \not\subset \mathfrak{p}_i$, whence $\mathfrak{a}_i + \mathfrak{p}_i = \mathfrak{o}$, since $\mathfrak{p}_i$ is a maximal prime ideal. It follows that there exists an element e_i of $\mathfrak{a}_i$ which is $\equiv 1 \pmod{\mathfrak{p}_i}$. Set $e_{i,n} = 1 - (1 - e_i^n)^n$; we have $e_{i,n} \in \mathfrak{a}_i^n$, $e_{i,n} \equiv 1 \pmod{\mathfrak{p}_i^n}$. The element $x = \sum_{i=1}^h a_i e_{i,n}$ is a solution of our system of congruences. If x' is any other solution, we have $(x' - x) \sum_{i=1}^h e_{i,n} \equiv 0 \pmod{(\mathfrak{p}_1 \cdots \mathfrak{p}_h)^n}$; by Lemma 1, $\sum_{i=1}^h e_{i,n}$ is a unit in $\mathfrak{o}$, whence $x' - x \equiv 0 \pmod{(\mathfrak{p}_1 \cdots \mathfrak{p}_h)^n}$.

PROPOSITION 1. *Let* $\mathfrak{o}$ *be a semi-local ring. If* $\mathfrak{b}$ *is an ideal* $\neq \mathfrak{o}$ *in* $\mathfrak{o}$, $\mathfrak{o}/\mathfrak{b}$ *is a semi-local ring. If* $\mathfrak{o}$ *is complete,* $\mathfrak{o}/\mathfrak{b}$ *is complete.*

Let $\mathfrak{p}_1, \cdots, \mathfrak{p}_{h_1}$ be the maximal prime ideals of $\mathfrak{o}$ which contain $\mathfrak{b}$, and let $\mathfrak{p}_{h_1+1}, \cdots, \mathfrak{p}_h$ be the other maximal prime ideals of $\mathfrak{o}$. The maximal prime ideals of $\mathfrak{o}/\mathfrak{b}$ are clearly the ideals $\mathfrak{p}_i/\mathfrak{b}$ $(1 \leqq i \leqq h_1)$, which proves that $\mathfrak{o}/\mathfrak{b}$ is semi-local. Assume that $\mathfrak{o}$ is complete, and let (u_n^*) be a convergent sequence of elements of $\mathfrak{o}/\mathfrak{b}$. The element $v_n^* = u_{n+1}^* - u_n^*$ is the residue class of an element v_n of $\mathfrak{o}$ which belongs to $(\mathfrak{p}_1 \cdots \mathfrak{p}_{h_1})^{m(n)}$, with $\lim_{n\to\infty} m(n) = \infty$. If $j > h_1$, we have $\mathfrak{b} + \mathfrak{p}_j = \mathfrak{o}$; it follows that $\mathfrak{p}_j$ contains an element $\bar{w}_j$ which is $\equiv 1 \pmod{\mathfrak{b}}$. We set $v_n' = v_n \cdot (\prod_{j>h_1}\bar{w}_j)^{m(n)}$; let u_1 be an element whose residue class modulo $\mathfrak{b}$ is u_1^*; we set $u_n = u_1 + \sum_{k=1}^{n-1} v_k'$; u_n is a representative of the residue class u_n^*, and the sequence (u_n) is convergent in $\mathfrak{o}$. Let u be its limit; it is clear that the sequence (u_n^*) converges in $\mathfrak{o}/\mathfrak{b}$ to the residue class u^* of u.

PROPOSITION 2. *Let* $\mathfrak{o}$ *be a complete local ring, and let* $\mathfrak{p}_1, \cdots, \mathfrak{p}_h$ *be the distinct maximal prime ideals of* $\mathfrak{o}$. *To every* $\mathfrak{p}_i$ $(1 \leqq i \leqq h)$ *there is associated an idempotent* ϵ_i *in* $\mathfrak{o}$, *with the following properties: if* $\mathfrak{o}\epsilon_i$ *is considered as a ring with the unit element* ϵ_i, *it is a complete semi-local ring whose only maximal prime ideal ideal is* $\mathfrak{p}_i\epsilon_i$; *we have* $\epsilon_i \in \mathfrak{p}_j$ *for* $j \neq i$; *the idempotents* $\epsilon_i, \cdots, \epsilon_h$ *are mutually orthogonal and their sum is* 1.

We make use of the elements $e_{i,n}$ which were constructed in the proof of Lemma 3. It follows immediately from Lemma 3 that $e_{i,n+1} - e_{i,n} \in (\mathfrak{p}_1 \cdots \mathfrak{p}_h)^n$; it follows that each one of the h sequences $(e_{i,n})$ is convergent, and has therefore

a limit ϵ_i in $\mathfrak{o}$. We have $e_{i,n}e_{j,n} \in (\mathfrak{p}_1 \cdots \mathfrak{p}_h)^n$ (for $i \neq j$), $1 - \sum_{i=1}^h e_{i,n} \in (\mathfrak{p}_1 \cdots \mathfrak{p}_h)^n$, whence $\epsilon_i\epsilon_j = 0$ for $i \neq j$ and $1 = \sum_{i=1}^h \epsilon_i$. Since $\epsilon_i = \lim_{n\to\infty} e_{i,n}$, $\epsilon_i - e_{i,n}$ is contained in $\mathfrak{p}_1 \cdots \mathfrak{p}_h$, and, *a fortiori*, in $\mathfrak{p}_j$ if n is large enough; if $j \neq i$, we have also $e_{i,n} \in \mathfrak{p}_j$, whence $\epsilon_i \in \mathfrak{p}_j$. Th ring $\mathfrak{o}\epsilon_i$ is isomorphic to $\mathfrak{o}/\mathfrak{o}(1 - \epsilon_i)$; Proposition 2 follows therefore from Proposition 1.

LEMMA 4.[4] *Let $\mathfrak{o}$ be a Noetherian ring which does not contain any zero divisor $\neq 0$ and let $\mathfrak{o}'$ be a ring which contains $\mathfrak{o}$ as a sub-ring and is a finite $\mathfrak{o}$-module. If $\mathfrak{a}$ is an ideal $\neq \mathfrak{o}$ in $\mathfrak{o}$, we have $\mathfrak{a}\mathfrak{o}' \neq \mathfrak{o}'$.*

Let us say that m elements $x_1, \cdots, x_m$ of $\mathfrak{o}'$ are linearly independent over $\mathfrak{o}$ if the conditions $\sum_{i=1}^m a_i x_i = 0$, $a_i \in \mathfrak{o}$ imply $a_i = 0$ $(1 \leqq i \leqq m)$. If $\mathfrak{o}' = \sum_{j=1}^M \mathfrak{o}y_j$, with $y_1 = 1$, we may extract from the set $\{y_1, \cdots, y_M\}$ a set of linearly independent elements $\{x_1, \cdots, x_m\}$ containing $x_1 = y_1 = 1$ and which is not properly contained in any other larger sub-set of $\{y_1, \cdots, y_M\}$ composed of linearly independent elements. It follows that there exists an element $c \neq 0$ in $\mathfrak{o}$ such that $c\mathfrak{o}' \subset \sum_{i=1}^m \mathfrak{o}x_i$, whence $c\mathfrak{a}^n\mathfrak{o}' \subset \sum_{i=1}^m \mathfrak{a}^n x_i$. Assume for a moment that $\mathfrak{a}\mathfrak{o}' = \mathfrak{o}'$, whence $1 \in \mathfrak{a}^n\mathfrak{o}'$ for every n. It follows that $c = \sum_{i=1}^m c_{i,n}x_i$ with $c_{i,n} \in \mathfrak{a}^n$ $(1 \leqq i \leqq m)$; since $x_1, \cdots, x_m$ are linearly independent over $\mathfrak{o}$ and $x_1 = 1$, we have $c = c_{1,n} \in \mathfrak{a}^n$ for all n, in contradiction with Lemma 1.

REMARK. The same argument proves that $\mathfrak{a}^n\mathfrak{o}' \cap \mathfrak{o} \subset \mathfrak{a}^n : \mathfrak{o}c$; we shall use this remark later.

LEMMA 5. *The result of Lemma 4 still holds if $\mathfrak{o}$ contains zero divisors.*

It is sufficient to consider the case where $\mathfrak{a}$ is a prime ideal $\mathfrak{p}$. The ideal $\mathfrak{p}$ contains some minimal prime divisor $\mathfrak{n}$ of the zero ideal. Let $\{0\} = \bigcap_i \mathfrak{v}_i$ be an irredundant representation of the zero ideal as an intersection of primary ideals, $\mathfrak{v}_1$ being the component of $\mathfrak{n}$; we have $\mathfrak{v}_1 : \mathfrak{n} \neq \mathfrak{v}_1$, whence $(\mathfrak{v}_1 : \mathfrak{n}) \cap (\bigcap_{i>1} \mathfrak{v}_i) \neq \{0\}$. If c is an element $\neq 0$ in the latter ideal, we have $c\mathfrak{n} = \{0\}$, $c \notin \mathfrak{v}_1$. Let $\mathfrak{n}'$ be the ideal $\mathfrak{n}\mathfrak{o}' \cap \mathfrak{o}$: we have $c\mathfrak{n}' = \{0\}$; if $\mathfrak{n}'$ were to contain an element ν' not contained in $\mathfrak{n}$, the equality $c\nu' = 0$ would imply $c \in \mathfrak{v}_1$, which is not the case. We have therefore $\mathfrak{n}' = \mathfrak{n}$. Since $\mathfrak{o}'$ is a finite $\mathfrak{o}$-module, $\mathfrak{o}'/\mathfrak{n}\mathfrak{o}'$ is a finite module over $\mathfrak{o}/\mathfrak{n}$. By Lemma 4, the ideal generated by $\mathfrak{p}/\mathfrak{n}$ in $\mathfrak{o}'/\mathfrak{n}\mathfrak{o}'$ is $\neq \mathfrak{o}'/\mathfrak{n}\mathfrak{o}'$, whence $\mathfrak{p}\mathfrak{o}' \neq \mathfrak{o}'$.

PROPOSITION 3. *Let $\mathfrak{o}$ be a semi-local ring, and let $\mathfrak{o}'$ be a ring which contains $\mathfrak{o}$ as a sub-ring and is a finite $\mathfrak{o}$-module. Then $\mathfrak{o}'$ is a semi-local ring. If $\mathfrak{o}$ is complete $\mathfrak{o}'$ is also complete.*

Let $\mathfrak{p}'$ be a maximal prime ideal in $\mathfrak{o}'$; if $\mathfrak{p} = \mathfrak{p}' \cap \mathfrak{o}$, the ring $\mathfrak{o}'/\mathfrak{p}'$ contains $\mathfrak{o}/\mathfrak{p}$ over which it is a finite module. But $\mathfrak{o}'/\mathfrak{p}'$ is a field; it follows immediately from Lemma 4 that $\mathfrak{o}/\mathfrak{p}$ cannot contain any ideal besides $\{0\}$ and $\mathfrak{o}/\mathfrak{p}$ itself; this means that $\mathfrak{o}/\mathfrak{p}$ is a field, i.e. that $\mathfrak{p}$ is a maximal prime ideal in $\mathfrak{o}$. Moreover, $\mathfrak{o}'/\mathfrak{p}\mathfrak{o}'$ is a finite module over $\mathfrak{o}/\mathfrak{p}$, i.e. is a hypercomplex system over $\mathfrak{o}/\mathfrak{p}$; it follows that there are only a finite number of prime ideals in $\mathfrak{o}'$ which contain $\mathfrak{p}$. We have therefore proved that $\mathfrak{o}'$ is a semi-local ring. Let $\mathfrak{a}$ be the product

---

[4] Lemma 4 is a generalization of a theorem of Krull (cf. Beitrage zur Arithmetik kommutativer Integritatsberiche, M. Z. 42, 1937, Satz 1, p. 749).

of the maximal prime ideals of $\mathfrak{o}$; every prime ideal in $\mathfrak{o}'$ which contains $\mathfrak{a}$ is maximal in $\mathfrak{o}'$, which proves that $\mathfrak{a}\mathfrak{o}'$ contains some power of the product $\mathfrak{a}'$ of all maximal prime ideals of $\mathfrak{o}'$. Assume now that $\mathfrak{o}$ is complete, and let (u'_n) be a convergent sequence of elements of $\mathfrak{o}'$; it follows from what we have said that $v'_n = u'_{n+1} - u'_n \in \mathfrak{a}^{m(n)}\mathfrak{o}'$, with $\lim_{n\to\infty} m(n) = \infty$. Let $y_1, \cdots, y_h$ be elements of $\mathfrak{o}'$ such that $\mathfrak{o}' = \sum_{j=1}^{h} \mathfrak{o} y_j$; we may write $v'_n = \sum_{i=1}^{h} v_{n,i} y_i$ with $v_{n,i} \in \mathfrak{a}^{m(n)}$. We set also $u'_1 = \sum_{i=1}^{h} u_{1,i} y_i$, $u_{n,i} = u_{1,i} + \sum_{k=1}^{n-1} v_{k,i}$. Each one of the sequences $(u_{n,i})$ is convergent in $\mathfrak{o}$; if we set $u_i = \lim_{n\to\infty} u_{n,i}$, the element $\sum_{i=1}^{h} u_i y_i - u'_n$ belongs to a power of $\mathfrak{a}\mathfrak{o}'$ whose exponent increases indefinitely with n, which proves that the sequence (u'_n) converges in $\mathfrak{o}'$ to the limit $\sum_{i=1}^{h} u_i y_i$.

LEMMA 6.[5] *Let $\mathfrak{o}$ be a semi-local ring, and let $\mathfrak{b}$ be an ideal in $\mathfrak{o}$. If $\mathfrak{a}$ is the product of the maximal prime ideals of $\mathfrak{o}$, we have $\bigcap_{n=1}^{\infty}(\mathfrak{b} + \mathfrak{a}^n) = \mathfrak{b}$.*

We know that $\mathfrak{o}/\mathfrak{b}$ is a semi-local ring and that $\mathfrak{a} + \mathfrak{b}/\mathfrak{b}$ is contained in the product of the maximal prime ideals of $\mathfrak{o}/\mathfrak{b}$ (cf. the proof of Proposition 1). It follows that $\bigcap_{n=1}^{\infty}(\mathfrak{a} + \mathfrak{b}/\mathfrak{b})^n = \bigcap_{n=1}^{\infty}(\mathfrak{a}^n + \mathfrak{b}/\mathfrak{b} = \{0\}$, which proves Lemma 6.

LEMMA 7. *Let $\mathfrak{o}$ be a complete semi-local ring, and let $(\mathfrak{b}_n)$ be a sequence of ideals in $\mathfrak{o}$ such that $\mathfrak{b}_{n+1} \subset \mathfrak{b}_n$ $(1 \leqq n < \infty)$ and $\bigcap_{n=1}^{\infty} \mathfrak{b}_n = \{0\}$. If $\mathfrak{p}_1, \cdots, \mathfrak{p}_h$ are the maximal prime ideals of $\mathfrak{o}$, we have $\mathfrak{b}_n \subset (\mathfrak{p}_1 \cdots \mathfrak{p}_h)^{m(n)}$, where $m(n)$ is an exponent which increases indefinitely with n.*

We set $\mathfrak{a} = \mathfrak{p}_1 \cdots \mathfrak{p}_h$; it follows immediately from Lemma 3 that $\mathfrak{o}/\mathfrak{a}^n$ is isomorphic to the direct product of the rings $\mathfrak{o}/\mathfrak{p}_i^n$ $(1 \leqq i \leqq h)$; since $\mathfrak{p}_i$ is the only prime ideal to contain $\mathfrak{p}_i^n$, $\mathfrak{p}_i^n$ is a primary ideal. It is well known that the descending chain condition for ideals holds in $\mathfrak{o}/\mathfrak{p}_i^n$; therefore it holds also in $\mathfrak{o}/\mathfrak{a}^n$. This being said, let h be any fixed integer; we set $\mathfrak{b}'_h = \bigcap_{n=1}^{\infty}(\mathfrak{a}_h + \mathfrak{b}_n)$. It follows from what we have just said that there exists an integer $N(h)$ such that $\mathfrak{a}^h + \mathfrak{b}_n = \mathfrak{b}'_h$ for every $n \geqq N(h)$; therefore we have $\mathfrak{b}'_h = \mathfrak{a}^h + \mathfrak{b}'_{h+1}$. Let b_h be any given element of $\mathfrak{b}'_h$; we define by induction an element $b_{h+n} \in \mathfrak{b}'_{h+n}$ in the following way: if b_{h+n} is already defined, we represent it in the form $a_{h+n} + b_{h+n+1}$, with $a_{h+n} \in \mathfrak{a}^{h+n}$, $b_{h+n+1} \in \mathfrak{b}'_{h+n+1}$, and this defines b_{h+n+1}. Our construction shows that the sequence (b_{h+n}) is convergent; let b be its limit. We have $b_{h+n} \equiv b_h \pmod{\mathfrak{a}^h}$ for every n, whence $b \equiv b_h \pmod{\mathfrak{a}^h}$; on the other hand, we have also $b \equiv b_{h+n} \pmod{\mathfrak{a}^{h+n}}$, and $b_{h+n} \in \bigcap_{m=1}^{\infty} (\mathfrak{a}^{h+n} + \mathfrak{b}_m)$, whence $b \in \mathfrak{a}^{h+n} + \mathfrak{b}_m$ for every m and n. By Lemma 6, we have $b \in \mathfrak{b}_m$ for every m, whence $b = 0$, $b_h \in \mathfrak{a}^h$, $\mathfrak{b}'_h = \mathfrak{a}^h$, $\mathfrak{b}_n \subset \mathfrak{a}^h$ for $n \geqq N(h)$, which proves Lemma 7.

PROPOSITION 4. *Let $\mathfrak{o}$ be a complete semi-local ring; we denote by $\mathfrak{p}_1, \cdots, \mathfrak{p}_h$ the maximal prime ideals of $\mathfrak{o}$ and by $\mathfrak{a}$ the ideal $\mathfrak{p}_1 \cdots \mathfrak{p}_h$. Let $\mathfrak{o}'$ be a ring which contains $\mathfrak{o}$ as a sub-ring. If $\bigcap_{n=1}^{\infty} \mathfrak{a}^n\mathfrak{o}' = \{0\}$, we have $\mathfrak{a}\mathfrak{o}' \cap \mathfrak{o} = \mathfrak{a}$, $\mathfrak{a}^n\mathfrak{o}' \cap \mathfrak{o} \subset \mathfrak{a}^{m(n)}$, where $m(n)$ is an exponent which increases indefinitely with n. If we assume furthermore that $\mathfrak{o}'/\mathfrak{a}\mathfrak{o}'$ is a finite module over $\mathfrak{o}/\mathfrak{a}$, $\mathfrak{o}'$ is a complete semi-local ring and is a finite $\mathfrak{o}$-module. Moreover, $\mathfrak{o}$ is a sub-space of $\mathfrak{o}'$.*

It follows immediately from Lemma 7 that $\mathfrak{a}^n\mathfrak{o}' \cap \mathfrak{o} \subset \mathfrak{a}^{m(n)}$, with $\lim_{n\to\infty}$

[5] This Lemma is due to Krull, Satz 2, l.c. note 3), p. 692.

$m(n) = \infty$. Therefore each $\mathfrak{p}_i$ contains $\mathfrak{a}^n\mathfrak{o}' \cap \mathfrak{o}$, and, a *fortiori*, $(\mathfrak{a}\mathfrak{o}' \cap \mathfrak{o})^n$ when n is sufficiently large; since $\mathfrak{p}_i$ is prime, we have $\mathfrak{a}\mathfrak{o}' \cap \mathfrak{o} \subset \bigcap_{i=1}^h \mathfrak{p}_i$. It follows immediately from Lemma 3 that $\bigcap_{i=1}^h \mathfrak{p}_i = \mathfrak{a}$, whence $\mathfrak{a}\mathfrak{o}' \cap \mathfrak{o} = \mathfrak{a}$. Assuming that $\mathfrak{o}'/\mathfrak{a}\mathfrak{o}'$ is a finite module over $\mathfrak{o}/\mathfrak{a}$, let $v_1, \cdots, v_d$ be elements of $\mathfrak{o}'$ such that $\mathfrak{o}'/\mathfrak{a}\mathfrak{o}' = \sum_{i=1}^d \mathfrak{o}/\mathfrak{a}.\ v_i^*$, where v_i^* is the residue class of v_i (mod. $\mathfrak{a}\mathfrak{o}'$). Every element of $\mathfrak{o}'$ is congruent (mod. $\mathfrak{a}\mathfrak{o}'$) to an element of $\sum_{i=1}^d \mathfrak{o}v_i$. Let also $\{a_1, \cdots, a_m\}$ be a set of generators of the ideal $\mathfrak{a}$, and let x be any element of $\mathfrak{o}'$; we construct by induction d sequences $(x_{i,n})_{n=0,1,\dots}$ of elements of $\mathfrak{o}$ such that $x \equiv \sum_{i=1}^m x_{i,n}v_i$ (mod. $\mathfrak{a}^n\mathfrak{o}'$). We set $x_{i,0} = 0$; assume that the elements $x_{i,n}$ have already been constructed; let $\xi_1, \cdots, \xi_N$ be all elements of $\mathfrak{o}$ which may be expressed as monomials of degree n in $a_1, \cdots, a_m$; they form a set of generators of $\mathfrak{a}^n\mathfrak{o}'$, whence $x - \sum_{i=1}^d x_{i,n}v_i = \sum_{k=1}^N y_k\xi_k$ with $y_k \in \mathfrak{o}'$; we have $y_k \equiv \sum_{i=1}^d y_{ki}v_i$ (mod. $\mathfrak{a}\mathfrak{o}'$) with $y_{ki} \in \mathfrak{o}$: we may take $x_{i,n+1} = x_{i,n} + \sum_{k=1}^N y_{ki}\xi_k$. At the same time, we see that each of the sequences $(x_{i,n})$ is convergent in $\mathfrak{o}$, and has therefore a limit x_i. We set $x' = \sum_{i=1}^d x_iv_i$; since $x_i \equiv x_{i,n}$ (mod. $\mathfrak{a}^n$), we have $x' - x \in \mathfrak{a}^n\mathfrak{o}'$ for every n, whence $x' = x$. We have proved that $\mathfrak{o}'$ is a finite $\mathfrak{o}$-module. Proposition 4 follows therefore from Proposition 3 above.

Let $\mathfrak{o}$ be a complete semi-local ring. Let A be any (finite or infinite) set of indices, and assume that we have assigned to every $\alpha \in A$ an element $u_\alpha \in \mathfrak{o}$, in such a way that, given any $k > 0$, we have $u_\alpha \in \mathfrak{a}^k$ for almost all α (i.e. for all but a finite number), where $\mathfrak{a}$ is the product of the maximal ideals of $\mathfrak{o}$. Let v_k be the sum of those elements u_α which are not in $\mathfrak{a}^k$; the sequence (v_k) has a limit v in $\mathfrak{o}$, and we shall set $v = \sum_{\alpha\in A} u_\alpha$. Such sums are associative; i.e., if A is represented as the union of a family of mutually disjoint sub-sets B_β, we have $\sum_{\alpha\in A} u_\alpha = \sum_\beta (\sum_{\alpha\in B_\beta} u_\alpha)$.

Let $\mathfrak{r}$ be a sub-ring of $\mathfrak{o}$; we shall denote by $\mathfrak{r}[[X_1, \cdots, X_m]]$ the ring of the formal power series in m letters $X_1, \cdots, X_m$ with coefficients in $\mathfrak{r}$. If F is any such power series, we may write $F = \sum_0^\infty F_k$, where F_k is a form of degree k. Let $a_1, \cdots, a_m$ be any m elements of $\mathfrak{a}$; the infinite sum $\sum_0^\infty F_k(a_1, \cdots, a_m)$ has a meaning in $\mathfrak{o}$; we shall set $F(a_1, \cdots, a_m) = \sum_0^\infty F_k(a_1, \cdots, a_m)$. It is clear that the mapping $F \to F(a_1, \cdots, a_m)$ is a homomorphism of $\mathfrak{r}[[X_1, \cdots, X_m]]$ into $\mathfrak{o}$. If this homomorphism turns out to be an isomorphism, we shall say that $a_1, \cdots, a_m$ are *analytically independent* over $\mathfrak{r}$.

We shall now prove that any semi-local ring can be imbedded in a certain complete semi-local ring. First, we prove

LEMMA 8. *Let $\mathfrak{o}$ be a Noetherian ring. The ring $\mathfrak{o}[[X_1, \cdots, X_r]]$ of the formal power series in r letters $X_1, \cdots, X_r$ with coefficients in $\mathfrak{o}$ is a Noetherian ring.*

Let $\mathfrak{x}$ be the ideal generated by $X_1, \cdots, X_r$. The elements of $\mathfrak{x}^n$ are the power series which do not contain any term of total degree $< n$. It follows immediately that $\bigcap_{n=1}^\infty \mathfrak{x}^n = \{0\}$. If F is any power-series $\neq 0$, and if $F \in \mathfrak{x}^d$, $F \notin \mathfrak{x}^{d+1}$ (with the convention that $\mathfrak{x}^0$ is the whole ring), we can find a homogeneous form F^* of degree d such that $F - F^* \in \mathfrak{x}^{d+1}$, and F^* is uniquely determined by this condition: F^* is called the beginning form of F. Let $\mathfrak{A}$ be an ideal $\neq \{0\}$ in $\mathfrak{o}[[X_1, \cdots, X_r]]$; the beginning forms of the elements $\neq 0$ in $\mathfrak{A}$

generate a homogeneous ideal $\mathfrak{A}^*$ in the ring $\mathfrak{o}[X_1, \cdots, X_r]$. The latter ring being Noetherian, $\mathfrak{A}^*$ has a finite set of generators $A_1^*, \cdots, A_m^*$; we may assume that each A_i^* is the beginning form of an element $A_i \in \mathfrak{A}$. Let d_i be the degree of the form A_i^*; if A^* is a form $\neq 0$ of degree d in $\mathfrak{A}^*$, we may write A^* in the form $\sum_i B_i A_i^*$, where each B_i is a form of degree $d - d_i (B_i = 0$ if $d < d_i)$; it follows immediately that A^* is the beginning form of the element $\sum_i B_i A_i \in \mathfrak{A}$.

Let C be any element of $\mathfrak{A}$; we shall define by induction m sequences $(C_{i,n})_{n=0,1,\ldots}$ of forms in the following way: we set $C_{i,0} = 0$ $(1 \leqq i \leqq m)$; if the elements $C_{i,k}$ have already been defined for $k < n$, we construct the element $C - \sum_{k=0}^{n-1} \sum_{i=0}^{m} C_{i,k} A_i$; if this element is 0, we set $C_{i,n} = 0$ for $1 \leqq i \leqq m$; if not, it has a beginning form of degree say e_n, and we may write this beginning form in the form $\sum_{i=1}^{m} C_{i,n} A_i^*$, where each $C_{i,n}$ is a form of degree $e_n - d_i$ $(C_{i,n} = 0$ if $d_i > e_n)$. If it happens for some n that all $C_{i,n}$ are 0, A clearly belongs to the ideal generated by $A_1, \cdots, A_m$ in $\mathfrak{o}[[X_1, \cdots, X_r]]$; if not, we have $e_{n+1} > e_n$ for all n, whence $\lim e_n = \infty$. It follows that $\sum_{n=0}^{\infty} C_{i,n}$ is a power series C_i and we have clearly $C = \sum_i A_i C_i$, which proves that $\mathfrak{A}$ is the ideal generated by $A_1, \cdots, A_m$. This shows that $\mathfrak{o}[[X_1, \cdots, X_r]]$ is a Noetherian ring.

Let (F_n) be any convergent sequence of elements of our ring such that $F_{n+1} - F_n$ belongs to a power of $\mathfrak{x}$ whose exponent increases indefinitely with n. If M is any monomial in $X_1, \cdots X_r$, the coefficient $a(M; F_n)$ of M in F_n remains constant when n is sufficiently large; let $a(M)$ be this constant value. The sum $\sum_M a(M)M$, extended over all monomials, represents a power series F; we shall set $F = \sum_1^{\infty} F_n$.

Let now $\mathfrak{o}$ be any semi-local ring; we denote by $\mathfrak{a}$ the product of the maximal prime ideals of $\mathfrak{o}$ and by $\{a_1, \cdots, a_m\}$ a set of generators of $\mathfrak{a}$. We construct with m letters $X_1, \cdots, X_m$ the ring $\mathfrak{o}[[X_1, \cdots, X_m]]$ of the power series with coefficients in $\mathfrak{o}$. If $F = \sum_0^{\infty} F_k$ is an element of this ring (with F_k homogeneous of degree k), we set $u_n(F) = \sum_{k=0}^{n} F_k(a_1, \cdots, a_m)$, and we denote by $\mathfrak{n}$ the set of the power-series F for which we have $\lim_{n \to \infty} u_n(F) = 0$. We see at once that $\mathfrak{n}$ is an ideal in $\mathfrak{o}[[X_1, \cdots, X_m]]$. We set $\bar{\mathfrak{o}} = \mathfrak{o}[[X_1, \cdots, X_m]]/\mathfrak{n}$. The ring $\mathfrak{o}$ is a sub-ring of $\mathfrak{o}[[X_1, \cdots, X_m]]$ and has obviously only 0 in common with $\mathfrak{n}$; we may therefore identify $\mathfrak{o}$ with the sub-ring of $\bar{\mathfrak{o}}$ upon which it is mapped isomorphically in the natural mapping of $\mathfrak{o}[[X_1, \cdots, X_m]]$ onto $\bar{\mathfrak{o}}$. Let $\mathfrak{x}$ be the ideal generated by $X_1, \cdots, X_m$ in $\mathfrak{o}[[X_1, \cdots, X_m]]$, and let $\bar{\mathfrak{a}}$ be the ideal $\mathfrak{x} + \mathfrak{n}/\mathfrak{n}$ in $\bar{\mathfrak{o}}$. We have clearly $X_i \equiv a_i \pmod{\mathfrak{n}}$ $(1 \leqq i \leqq m)$, whence $\bar{\mathfrak{a}} = \mathfrak{a}\bar{\mathfrak{o}}$. We shall prove that $\bar{\mathfrak{a}}^n \cap \mathfrak{o} = \mathfrak{a}^n$. Let y be any element of $\bar{\mathfrak{a}}^n \cap \mathfrak{o}$; y is the residue class modulo $\mathfrak{n}$ of a power series F such that $F \in \mathfrak{x}^n$, $F - y \in \mathfrak{n}$; since $\lim_{k \to \infty} (u_k(F) - y) = 0$, there exists an index $N > n$ such that $y \equiv u_N(F)$ $(\text{mod. } \mathfrak{a}^n)$; since $F \in \mathfrak{x}^n$, we have $u_N(F) = \sum_{k=n}^{N} F_k(a_1, \cdots, a_m) \in \mathfrak{a}^n$, whence $y \in \mathfrak{a}^n$, which proves our assertion.

It is clear that, if $u \in \bar{\mathfrak{o}}$ and n is given > 0, there always exists an element $u' \in \mathfrak{o}$ such that $u - u' \in \bar{\mathfrak{a}}^n$. Since $\bar{\mathfrak{a}} \cap \mathfrak{o} = \mathfrak{a}$, it follows that $\bar{\mathfrak{o}}/\bar{\mathfrak{a}}$ is isomorphic with $\mathfrak{o}/\mathfrak{a}$. If $\mathfrak{p}$ is any maximal prime ideal in $\mathfrak{o}$, we have $\bar{\mathfrak{a}} \subset \mathfrak{p}\bar{\mathfrak{o}}$, and $\bar{\mathfrak{o}}/\mathfrak{p}\bar{\mathfrak{o}}$ is

isomorphic with $\mathfrak{o}/\mathfrak{p}$; it follows that $\mathfrak{p}\bar{\mathfrak{o}}$ is a maximal prime ideal in $\bar{\mathfrak{o}}$. Any element u of $\bar{\mathfrak{o}}$ which is $\equiv 1 \pmod{\bar{\mathfrak{a}}}$ is a unit; in fact, u is the residue class (mod. $\mathfrak{n}$) of a power series of the form $1 + X$, with $X \in \mathfrak{x}$; such a series has an inverse in $\mathfrak{o}[[X_1, \cdots, X_m]]$, which can be represented by the series $1 - X + X^2 - \cdots + (-1)^n X^n + \cdots$; it follows that u has an inverse in $\mathfrak{o}$. If $\bar{\mathfrak{p}}$ is any maximal prime ideal in $\mathfrak{o}$, we have therefore $1 \notin \bar{\mathfrak{p}} + \bar{\mathfrak{a}}$, whence $\bar{\mathfrak{a}} \subset \bar{\mathfrak{p}}$, and it follows that $\bar{\mathfrak{p}}$ is of the form $\mathfrak{p}\bar{\mathfrak{o}}$, $\mathfrak{p}$ being some maximal prime ideal in $\mathfrak{o}$. We know that $\bar{\mathfrak{o}}$ is Noetherian by Lemma 8; it follows that $\bar{\mathfrak{o}}$ is a semi-local ring. If (u_n) is a sequence of elements of $\bar{\mathfrak{o}}$ which is convergent in $\bar{\mathfrak{o}}$, we have $u_{n+1} - u_n \in \bar{\mathfrak{a}}^{m(n)}$, with $\lim_{n\to\infty} m(n) = \infty$. It follows that $u_{n+1} - u_n$ is the residue class modulo $\mathfrak{n}$ of a power series $V_n \in \mathfrak{x}^{m(n)}$. The sum $\sum_1^\infty V_n$ represents a power series V; if v is the residue class of V, we have clearly $u_1 + v = \lim_{n\to\infty} u_n$, which proves that $\bar{\mathfrak{o}}$ is complete.

Let now $\mathfrak{o}^*$ be a complete semi-local ring which contains $\mathfrak{o}$ as a sub-ring and is such that $\mathfrak{a}$ is contained in the product $\mathfrak{a}^*$ of the maximal prime ideals of $\mathfrak{o}^*$. Let F be any element of $\mathfrak{o}[[X_1, \cdots, X_m]]$; the sequence $(u_n(F))$ is convergent in $\mathfrak{o}$, and therefore also in $\mathfrak{o}^*$; as such it has a limit $\theta_1(F)$ in $\mathfrak{o}^*$; θ_1 is clearly a homomorphism of $\mathfrak{o}[[X_1, \cdots, X_m]]$ into $\mathfrak{o}^*$ and maps $\mathfrak{n}$ onto $\{0\}$. It follows that θ_1 defines in a natural way a homomorphism θ of $\bar{\mathfrak{o}}$ into $\mathfrak{o}^*$ which maps every element of $\mathfrak{o}$ upon itself. The ring $\theta(\bar{\mathfrak{o}})$ is contained in the adherence of $\mathfrak{o}$, considered as a sub-set of $\mathfrak{o}^*$; on the other hand, $\theta(\bar{\mathfrak{o}})$ is a complete semi-local ring; it follows immediately from Lemma 7 that $\theta(\bar{\mathfrak{o}})$ is a sub-space of $\mathfrak{o}^*$. Since $\theta(\bar{\mathfrak{o}})$ is complete, it must be a closed sub-space of $\mathfrak{o}^*\cdot$ $\theta(\bar{\mathfrak{o}})$ is therefore the adherence of $\mathfrak{o}$ in $\mathfrak{o}^*$.

If $\mathfrak{o}$ is everywhere dense in $\mathfrak{o}^*$, we have $\theta(\bar{\mathfrak{o}}) = \mathfrak{o}^*$: We may observe also that, in this case, the condition $\mathfrak{a} \subset \mathfrak{a}^*$ is automatically verified. In fact, $\mathfrak{o}^*/\mathfrak{a}^*$ is isomorphic to $\mathfrak{o}/\mathfrak{a}^* \cap \mathfrak{o}$, which proves that every prime ideal of $\mathfrak{o}$ containing $\mathfrak{a}^* \cap \mathfrak{o}$ is maximal in $\mathfrak{o}$, and that $\mathfrak{a}^* \cap \mathfrak{o}$ is the product of these prime ideals, whence $\mathfrak{a} \subset \mathfrak{a}^* \cap \mathfrak{o}$.

If $\mathfrak{o}$ is a sub-space of $\mathfrak{o}^*$, it is clear that the elements of $\mathfrak{n}$ are the only ones to be mapped on 0 under θ; in this case, θ is therefore an isomorphism. We have proved

THEOREM 1.[6] *Let $\mathfrak{o}$ be a semi-local ring. There exists a complete semi-local ring $\mathfrak{o}$ containing $\bar{\mathfrak{o}}$ as a sub-ring and a sub-space in which $\bar{\mathfrak{o}}$ is everywhere dense. If $\mathfrak{o}_1$ is another ring with the same properties, there exists an isomorphism of $\bar{\mathfrak{o}}$ with $\mathfrak{o}_1$ which maps every element of $\mathfrak{o}$ upon itself. If $\mathfrak{o}^*$ is any complete local ring containing $\mathfrak{o}$ as a sub-ring, and if the product of the maximal prime ideals of $\mathfrak{o}$ is contained in the product of the maximal prime ideals of $\mathfrak{o}^*$, the adherence of $\mathfrak{o}$ in $\mathfrak{o}^*$ is a complete local ring which is a homomorphic image of $\bar{\mathfrak{o}}$. If $\mathfrak{o}$ is a sub-space of $\mathfrak{o}^*$, the adherence of $\mathfrak{o}$ in $\mathfrak{o}^*$ is isomorphic with $\bar{\mathfrak{o}}$.*

DEFINITION 3. *The ring $\bar{\mathfrak{o}}$ which is described in Theorem 1 is called the completion of the semi-local ring $\mathfrak{o}$.*

[6] Theorem 1 is a precision of Satz 14 of Krull, l.c. note 3), p. 692.

PROPOSITION 5. *If $\mathfrak{b}$ is an ideal in a semi-local ring $\mathfrak{o}$, and if $\bar{\mathfrak{o}}$ is the completion of $\mathfrak{o}$, we have $\mathfrak{b}\bar{\mathfrak{o}} \cap \mathfrak{o} = \mathfrak{b}$. The ring $\bar{\mathfrak{o}}/\mathfrak{b}\bar{\mathfrak{o}}$ is the completion of $\mathfrak{o}/\mathfrak{b}$.*

Let $\{b_1, \cdots, b_m)\}$ be a set of generators of $\mathfrak{b}$; if $u \in \mathfrak{b}\bar{\mathfrak{o}} \cap \mathfrak{o}$, we have $u = \sum_{i=1}^{m} b_i u_i$, $u_i \in \bar{\mathfrak{o}}$; if n is any given exponent, we can find elements $u_{i,n} \in \mathfrak{o}$ such that $u_i \equiv u_{i,n}$ (mod. $\mathfrak{a}^n\bar{\mathfrak{o}}$), where $\mathfrak{a}$ is the product of the maximal prime ideals of $\mathfrak{o}$. We have $b - \sum_i b_i u_{i,n} \in \mathfrak{a}^n\bar{\mathfrak{o}} \cap \mathfrak{o} = \mathfrak{a}^n$ (cf. the proof of Theorem 1), and $\sum_i b_i u_{i,n} \in \mathfrak{b}$, whence $u \in \mathfrak{b} + \mathfrak{a}^n$; by Lemma 6, we have $u \in \mathfrak{b}$, whence $\mathfrak{b}\bar{\mathfrak{o}} \cap \mathfrak{o} = \mathfrak{b}$. It follows that $\bar{\mathfrak{o}}/\mathfrak{b}\bar{\mathfrak{o}}$ contains $\mathfrak{o}/\mathfrak{b}$ as a sub-ring; this sub-ring is clearly everywhere dense. The product of the maximal prime ideals of $\mathfrak{o}/\mathfrak{b}$ is $\mathfrak{a} + \mathfrak{b}/\mathfrak{b}$; the product of the maximal prime ideals of $\bar{\mathfrak{o}}/\mathfrak{b}\bar{\mathfrak{o}}$ is $\mathfrak{a} + \mathfrak{b}\bar{\mathfrak{o}}/\mathfrak{b}\bar{\mathfrak{o}}$. Let u^* be any element of $(\mathfrak{a} + \mathfrak{b}\bar{\mathfrak{o}}/\mathfrak{b}\bar{\mathfrak{o}})^n \cap \mathfrak{o}/\mathfrak{b}$; u^* is the residue class modulo $\mathfrak{b}\bar{\mathfrak{o}}$ of an element $u \in \mathfrak{a}^n\bar{\mathfrak{o}}$ and is also the residue class modulo $\mathfrak{b}$ of an element $u' \in \mathfrak{o}$. Since $\mathfrak{b}\bar{\mathfrak{o}} \subset \mathfrak{b} + \mathfrak{a}^n\bar{\mathfrak{o}}$, there exists an element $b \in \mathfrak{b}$ such that $u' - b \in \mathfrak{a}^n\bar{\mathfrak{o}} \cap \mathfrak{o} = \mathfrak{a}^n$ whence $u^* \in (\mathfrak{a} + \mathfrak{b}/\mathfrak{b})^n$. We have proved that $(\mathfrak{a} + \mathfrak{b}\bar{\mathfrak{o}}/\mathfrak{b}\bar{\mathfrak{o}})^n \cap \mathfrak{o}/\mathfrak{b} \doteq (\mathfrak{a} + \mathfrak{b}/\mathfrak{b})^n$, which shows that $\mathfrak{o}/\mathfrak{b}$ is a sub-space of $\bar{\mathfrak{o}}/\mathfrak{b}\bar{\mathfrak{o}}$: Proposition 5 is proved.

PROPOSITION 6. *Let $\bar{\mathfrak{o}}$ be the completion of a semi-local ring $\mathfrak{o}$. If an element $u \in \mathfrak{o}$ is not a zero divisor in $\mathfrak{o}$, it is not a zero divisor in $\bar{\mathfrak{o}}$.*

Assume that $uv = 0$, $v \in \bar{\mathfrak{o}}$. If $\mathfrak{a}$ is the product of the maximal prime ideals in $\mathfrak{o}$, we can find for every $n > 0$ an element $v_n \in \mathfrak{o}$ such that $v - v_n \in \mathfrak{a}^n\bar{\mathfrak{o}}$. We have $uv_n \in u\mathfrak{a}^n\bar{\mathfrak{o}} \cap \mathfrak{o} = u\mathfrak{a}^n$. Since u is not a zero divisor in $\mathfrak{o}$, it follows that $v_n \in \mathfrak{a}^n$, whence $v = 0$, which proves Proposition 6.

LEMMA 9. *Let $\mathfrak{a}$ be the product of the maximal prime ideals in a semi-local ring $\mathfrak{o}$, and let c be an element of $\mathfrak{o}$ which is not a zero divisor. Then $\mathfrak{a}^n : \mathfrak{o}c$ is contained in a power of $\mathfrak{a}$ whose exponent increases indefinitely with n.*

Let $\bar{\mathfrak{o}}$ be the completion of $\mathfrak{o}$. If $u \in \mathfrak{a}^n\bar{\mathfrak{o}} : \bar{\mathfrak{o}}c$ for very $n > 0$, we have $cu \in \bigcap_{n=1}^{\infty} \mathfrak{a}^n\bar{\mathfrak{o}} = \{0\}$, whence $u = 0$ by Proposition 6. By Lemma 7, we have $\mathfrak{a}^n\bar{\mathfrak{o}} : \bar{\mathfrak{o}}c \subset \mathfrak{a}^{m(n)}\bar{\mathfrak{o}}$, with $\lim_{n\to\infty} m(n) = \infty$. Lemma 9 follows if we observe that $\mathfrak{a}^n : \mathfrak{o}c \subset \mathfrak{a}^n\bar{\mathfrak{o}} : \bar{\mathfrak{o}}c \cap \mathfrak{o}$.

PROPOSITION 7. *Let $\mathfrak{o}$, $\mathfrak{o}'$ be two semi-local rings on which we make the following assumptions: $\mathfrak{o}'$ contains $\mathfrak{o}$ as a sub-ring and is a finite $\mathfrak{o}$-module; no element $\neq 0$ in $\mathfrak{o}$ is a zero divisor in $\mathfrak{o}'$. Let $\bar{\mathfrak{o}}$, $\bar{\mathfrak{o}}'$ be the completions of $\mathfrak{o}$, $\mathfrak{o}'$ respectively. Then $\bar{\mathfrak{o}}$ is a sub-ring of $\bar{\mathfrak{o}}'$ (up to an isomorphism); if $\mathfrak{o}' = \sum_{i=1}^{m} \mathfrak{o}y_i$, we have $\bar{\mathfrak{o}}' = \sum_{i=1}^{m} \bar{\mathfrak{o}}y_i$. If m elements $x_1, \cdots, x_m$ of $\mathfrak{o}'$ are linearly independent over $\mathfrak{o}$, they are also linearly independent over $\bar{\mathfrak{o}}$. If an element of $\bar{\mathfrak{o}}$ is a zero divisor in $\bar{\mathfrak{o}}'$, it is already a zero divisor in $\bar{\mathfrak{o}}$.*

It follows from Lemma 9 and from the remark which follows the proof of Lemma 4 that $\mathfrak{o}$ is a sub-space of $\mathfrak{o}'$, and therefore also of $\bar{\mathfrak{o}}'$. By Theorem 1, we may assume that $\bar{\mathfrak{o}} \subset \bar{\mathfrak{o}}'$. Set $\bar{\mathfrak{o}}_1' = \sum_{i=1}^{m} \mathfrak{o}y_i$; $\bar{\mathfrak{o}}_1'$ is a finite $\bar{\mathfrak{o}}$-module, and is therefore a complete semi-local ring. If $\mathfrak{a}$ is the product of the maximal prime ideals of $\mathfrak{o}$, $\mathfrak{a}\bar{\mathfrak{o}}'$ is contained in every maximal prime ideal of $\mathfrak{o}'$, and therefore $\mathfrak{a}\bar{\mathfrak{o}}'$ is contained in every maximal prime ideal of $\bar{\mathfrak{o}}'$ whence $\bigcap_{n=1}^{\infty} \mathfrak{a}^n\bar{\mathfrak{o}}' = \{0\}$. It follows from Lemma 7 that $\bar{\mathfrak{o}}_1'$ is a sub-space of $\bar{\mathfrak{o}}'$; since $\bar{\mathfrak{o}}_1'$ is complete, it is a closed sub-set of $\bar{\mathfrak{o}}'$; since $\mathfrak{o}' \subset \bar{\mathfrak{o}}_1'$, we have $\bar{\mathfrak{o}}' = \bar{\mathfrak{o}}_1'$.

If the elements $x_1, \cdots, x_m$ are linearly independent over $\mathfrak{o}$, we may extend

the set $\{x_1, \cdots, x_m\}$ to a maximal set of linearly independent elements (for instance by adjoining some of the elements y_j). We may therefore assume without loss of generality that the set $\{x_1, \cdots x_m\}$ is already a maximal set of linearly independent elements. It follows that there exists an element $c \neq 0$ in $\mathfrak{o}$ such that $c\mathfrak{o}' \subset \sum_{i=1}^m \mathfrak{o}x_i$. This being said, assume that $\sum_{i=1}^m u_ix_i = 0$ ($u_i \in \bar{\mathfrak{o}}$). For every $n > 0$, we can find elements $u_{i,n} \in \mathfrak{o}$ such that $u_i - u_{i,n} \in \mathfrak{a}^n\bar{\mathfrak{o}}$, whence $\sum_{i=1}^n u_{i,n}x_i \in \sum_{i=1}^n \mathfrak{a}^n\bar{\mathfrak{o}}'x_i \cap \mathfrak{o}' \subset \mathfrak{a}^n\mathfrak{o}'$. It follows that $\sum_{i=1}^m cu_{i,n}x_i \in \sum_{i=1}^m \mathfrak{a}^n x_i$, $\sum_{i=1}^m cu_{i,n}x_i = \sum_{i=1}^m a_{i,n}x_i$, $a_{i,n} \in \mathfrak{a}^n$. The elements x_i being linearly independent over $\mathfrak{o}$, we have $cu_{i,n} = a_{i,n} \in \mathfrak{a}^n$, $u_{i,n} \in \mathfrak{a}^n:\mathfrak{o}c \subset \mathfrak{a}^{m(n)}$ with $\lim_{n\to\infty} m(n) = \infty$ (by Lemma 9). It follows that $u_i = \lim_{n\to\infty} u_{i,n} = 0$, which proves that $x_1, \cdots, x_m$ are linearly independent over $\bar{\mathfrak{o}}$.

Let u be an element of $\bar{\mathfrak{o}}$, and assume that $uv = 0$, $v \in \bar{\mathfrak{o}}'$. Using the same notations as above, we may write $cv = \sum_{i=1}^m u_ix_i$, $u_i \in \bar{\mathfrak{o}}$, whence $\sum_{i=1}^m uu_ix_i = 0$, $uu_i = 0$ $(1 \leqq i \leqq m)$. If u is not a zero divisor in $\bar{\mathfrak{o}}$, we have $u_i = 0$ $(i \leqq i \leqq m)$, whence $cv = 0$ and $v = 0$ by Proposition 6.

Proposition 8. *Let $\mathfrak{o}$ be a semi-local ring which does not contain any zero divisor $\neq 0$, and let $\mathfrak{p}_1, \cdots, \mathfrak{p}_h$ be the maximal prime ideals of $\mathfrak{o}$. Let $\mathfrak{o}_i$ be the quotient ring of $\mathfrak{p}_i$ with respect to $\mathfrak{o}$. The completion of $\mathfrak{o}$ is isomorphic to the direct product of the completions of the rings $\mathfrak{o}_i$.*

We use, for the ring $\bar{\mathfrak{o}}$, the notations of Proposition 2. It follows from Proposition 6 that $\mathfrak{o}\epsilon_i$ is a ring which is isomorphic with $\mathfrak{o}$. Let a be an element of $\mathfrak{o}$ which does not belong to $\mathfrak{p}_i$; since $\mathfrak{p}_i$ is maximal in $\mathfrak{o}$, there exists an element $b \in \mathfrak{o}$ such that $ab \equiv 1$ (mod. $\mathfrak{p}_i$); it follows that $ab\epsilon_i$ has an inverse in $\bar{\mathfrak{o}}\epsilon_i$; this means that $\bar{\mathfrak{o}}\epsilon_i$ contains the quotient ring $\mathfrak{o}_i\epsilon_i$ of $\mathfrak{p}_i\epsilon_i$ with respect to $\mathfrak{o}\epsilon_i$. This quotient ring is clearly everywhere dense in $\bar{\mathfrak{o}}\epsilon_i$. Let a be an element of $\mathfrak{o}$ such that $a\epsilon_i \in \mathfrak{p}_i^n\bar{\mathfrak{o}}\epsilon_i$; we have $a = a\epsilon_i + a(1 - \epsilon_i) \in \mathfrak{p}_i^n\bar{\mathfrak{o}}$, since $1 - \epsilon_i \in \mathfrak{p}_i^n\bar{\mathfrak{o}}$ for every n; it follows that $a\epsilon_i \in \mathfrak{p}_i^n\epsilon_i$. The ideal $\mathfrak{p}_i^n\bar{\mathfrak{o}} \cap \mathfrak{o}_i\epsilon_i$ is an ideal of $\mathfrak{o}_i\epsilon_i$ whose intersection with $\mathfrak{o}\epsilon_i$ is $\mathfrak{p}_i^n\epsilon_i$; by Lemma 2, §I, p. 691, this ideal is $\mathfrak{p}_i^n\mathfrak{o}_i\epsilon_i$, which proves that $\mathfrak{o}_i\epsilon_i$ is a sub-space of $\bar{\mathfrak{o}}\epsilon_i$; $\bar{\mathfrak{o}}\epsilon_i$ is therefore the completion of $\mathfrak{o}_i\epsilon_i$, which proves Proposition 8.

§III. Local Rings

The notion of local ring has been introduced by Krull.[7] He has defined a local ring as being a Noetherian ring in which the non-units form an ideal. The local rings which we shall consider in this paper are however of a more restricted type. Before defining them, we make the following observation: if a ring $\mathfrak{o}$ contains a field K as a sub-field, and if $\mathfrak{m}$ is an ideal in $\mathfrak{o}$, the field K is mapped isomorphically under the natural mapping of $\mathfrak{o}$ onto $\mathfrak{o}/\mathfrak{m}$, and therefore K may be identified with a sub-field of the ring $\mathfrak{o}/\mathfrak{m}$. This being said, we set up the following definition:

Definition 1. *A local ring is a Noetherian ring $\mathfrak{o}$ which satisfies the following conditions:* a) *the non-units in $\mathfrak{o}$ form an ideal* $\mathfrak{m}$; b) *the ring $\mathfrak{o}$ contains a sub-*

[7] Cf. Krull, l.c. note 3), p. 692.

field K with infinitely many elements[8] *such that* $\mathfrak{o}/\mathfrak{m}$ *is a finite K-module. The ring* $\mathfrak{o}/\mathfrak{m}$, *which is necessarily a field, is called the field of residues of* $\mathfrak{o}$. *Any sub-field K of* $\mathfrak{o}$ *which satisfies condition 2) is called a basic field of* $\mathfrak{o}$.

DEFINITION 2. *Let* $\mathfrak{o}$ *be a local ring, and let* $\mathfrak{m}$ *be the ideal of non-units in* $\mathfrak{o}$. *We shall say that* $\mathfrak{o}$ *is of dimension r if it is possible to find a set of r elements of* $\mathfrak{o}$ *which generates an ideal which is primary for* $\mathfrak{m}$, *but if it is impossible to find a set of less than r elements with the same property. Any set of r elements which generates an ideal which is primary for* $\mathfrak{m}$ *is then called a system of parameters in* $\mathfrak{o}$.[9]

We include in this definition the case $r = 0$ by making the convention that the ideal generated by the empty set is the zero-ideal.

Since $\mathfrak{m}$ is a maximal prime ideal, the condition for an ideal of being primary for $\mathfrak{m}$ is clearly equivalent to the condition of containing some power of $\mathfrak{m}$. Therefore, a local ring of dimension 0 is a local ring in which every non-unit is nilpotent.

PROPOSITION 1. *If a ring* $\mathfrak{o}'$ *is a homomorphic image of a local ring* $\mathfrak{o}$, *it is a local ring and we have* $\dim \mathfrak{o}' \leqq \dim \mathfrak{o}$. *In the case where* $\dim \mathfrak{o}' = \dim \mathfrak{o}$, *the images in* $\mathfrak{o}'$ *of the elements of a system of parameters in* $\mathfrak{o}$ *form a system of parameters in* $\mathfrak{o}'$.

By Proposition 1, §II, p. 693, we know that $\mathfrak{o}'$ is a semi-local ring; the only maximal prime ideal of $\mathfrak{o}'$ is clearly the image of the ideal of non units in $\mathfrak{o}$. The image of a basic field of $\mathfrak{o}$ is a basic field of $\mathfrak{o}'$, which proves that $\mathfrak{o}'$ is a local ring. If $x_1, \cdots, x_r$ are elements of $\mathfrak{o}$ which generate a primary ideal for the ideal $\mathfrak{m}$ of non units, their images in $\mathfrak{o}'$ generate an ideal which is primary for the ideal of non units of $\mathfrak{o}'$, which proves Proposition 1.

Since a local ring is also semi-local, we may use the notions of "complete", or of "completion", (and more generally all the topological notions), in connection with a local ring.

PROPOSITION 2. *The completion* $\bar{\mathfrak{o}}$ *of local ring* $\mathfrak{o}$ *of dimension r is a local ring of dimension r. Any system of parameters in* $\mathfrak{o}$ *is also a system of parameters in* $\bar{\mathfrak{o}}$.

If $\mathfrak{m}$ is the ideal of non units in $\mathfrak{o}$, we know that $\bar{\mathfrak{o}}$ is a semi-local ring in which every maximal prime ideal contains $\mathfrak{m}\bar{\mathfrak{o}}$; since $\mathfrak{o}$ is everywhere dense in $\bar{\mathfrak{o}}$, $\bar{\mathfrak{o}}/\mathfrak{m}\bar{\mathfrak{o}}$ is isomorphic with $\mathfrak{o}/\mathfrak{m}$, i.e. is a field. It follows that $\mathfrak{m}\bar{\mathfrak{o}}$ is prime and that a basic field of $\mathfrak{o}$ is also a basic field of $\bar{\mathfrak{o}}$:$\bar{\mathfrak{o}}$ is a local ring. The elements of a system of parameters in $\mathfrak{o}$ generate in $\bar{\mathfrak{o}}$ an ideal which is primary for $\mathfrak{m}\bar{\mathfrak{o}}$. Let conversely $\{\bar{x}_1, \cdots, \bar{x}_s\}$ be a system of parameters in $\bar{\mathfrak{o}}$, and assume that $\bar{\mathfrak{q}} = \sum_{i=1}^{s} \bar{\mathfrak{o}}\bar{x}_i$ contains $\mathfrak{m}^e\bar{\mathfrak{o}}$. We can find s elements $x_1, \cdots, x_s$ in $\mathfrak{o}$ such that $x_i \equiv \bar{x}_i$ (mod. $\mathfrak{m}^{e+1}\bar{\mathfrak{o}}$) $(1 \leqq i \leqq s)$. If we set $\mathfrak{q} = \sum_{i=1}^{s} \bar{\mathfrak{o}}x_i$, we have $\bar{\mathfrak{q}} \subset \mathfrak{q} + \mathfrak{m}^{e+1}\bar{\mathfrak{o}}$, and, by induction, $\mathfrak{m}^e\bar{\mathfrak{o}} \subset \mathfrak{q} + \mathfrak{m}^{e+n}\bar{\mathfrak{o}}$ for every n. By Lemma 6, §II, p. 695, it follows that $\mathfrak{m}^e\bar{\mathfrak{o}} \subset \mathfrak{q}$, $\mathfrak{m}^e \subset \mathfrak{q} \cap \mathfrak{o} = \sum_{i=1}^{s} \mathfrak{o}x_i$, which proves that $\mathfrak{o}$ is of dimension $\leqq s$. Proposition 2 is proved.

PROPOSITION 3. *Let* $\mathfrak{o}$ *be a complete local ring and let* $\mathfrak{m}$ *be the ideal of non units*

[8] This restriction is not really necessary, but makes some of the arguments simpler.

[9] This definition of the dimension is equivalent to the one given by Krull, l.c. note 3), p. 692. (Cf. The appendix at the end of this paper.)

in $\mathfrak{o}$; if there exists a basic field K of $\mathfrak{o}$ such that $\mathfrak{o}/\mathfrak{m}$ is separable over K, K is contained in a sub-field K' of $\mathfrak{o}$ which is a complete system of representatives for the residue classes of $\mathfrak{o}$ modulo $\mathfrak{m}$.[10]

Let ζ be an element of $\mathfrak{o}$ whose residue class modulo $\mathfrak{m}$ generates $\mathfrak{o}/\mathfrak{m}$ over K. There exists an irreducible polynomial $F(Z)$ with coefficients in K such that $F(\zeta) \equiv 0 \pmod{\mathfrak{m}}$, $dF/dZ(\zeta) \not\equiv 0 \pmod{\mathfrak{m}}$. It follows that $dF/dZ(\zeta)$ has an inverse a in $\mathfrak{o}$. We shall construct by induction a sequence of elements $\zeta_k \in \mathfrak{o}$ such that $F(\zeta_k) \equiv 0 \pmod{\mathfrak{m}^k}$. We set $\zeta_1 = \zeta$; if ζ_k is already constructed, we set $\zeta_{k+1} = \zeta_k - aF(\zeta_k)$; it follows immediately from the Taylor expansion theorem that $F(\zeta_{k+1}) \equiv 0 \pmod{\mathfrak{m}^{2k}}$. Moreover our construction shows that the sequence (ζ_k) is convergent; if ζ_∞ is its limit, we have $F(\zeta_\infty) = 0$. Since ζ_∞ is algebraic over K, we have $K(\zeta_\infty) = K[\zeta_\infty] \subset \mathfrak{o}$, and $K(\zeta_\infty)$ is clearly a complete system of representatives for the classes modulo $\mathfrak{m}$.

PROPOSITION 4. *Let $\mathfrak{o}$ be a complete local ring, and let $\{x_1, \cdots, x_r\}$ be a system of parameters in $\mathfrak{o}$. If K is a basic field of $\mathfrak{o}$, the elements $x_1, \cdots, x_r$ are analytically independent over K.*

Proposition 3 will follow immediately from

LEMMA 1. *Let $\mathfrak{o}$ be a local ring; let $\{x_1, \cdots, x_r\}$ be a system of parameters in $\mathfrak{o}$; let F be a form of degree d in r letters with coefficients in $\mathfrak{o}$; if one at least of the coefficients of F does not belong to the ideal of non units $\mathfrak{m}$ of $\mathfrak{o}$, we have $F(x_1, \cdots, x_r) \notin \mathfrak{x}^{d+1}$, where $\mathfrak{x} = \sum_{i=1}^{r} \mathfrak{o}x_i$.*[11]

Assume for a moment that $F(x_1, \cdots, x_r) \in \mathfrak{x}^{d+1}$; let F^* be the form with coefficients in $\mathfrak{o}/\mathfrak{m}$ which is deduced from F by reduction of the coefficients modulo $\mathfrak{m}$; since $\mathfrak{o}/\mathfrak{m}$ contains infinitely many elements, we can find $r - 1$ elements $a_1^*, \cdots, a_{r-1}^*$ of this field such that $F^*(a_1^*, \cdots, a_{r-1}^*, 1) \neq 0$; we select for each i $(1 \leqq i \leqq r - 1)$ a representative $a_i \in \mathfrak{o}$ of the class a_i^*, and we set $x_i' = x_i - a_i x_r$. Let $\mathfrak{x}'$ be the ideal $\sum_{i=1}^{r-1} \mathfrak{o}x_i'$. We have $F(x_1, \cdots, x_r) \equiv F(a_1, \cdots, a_{r-1}, 1)x_r^d \pmod{\mathfrak{x}'}$ and $F(a_1, \cdots, a_{r-1}, 1)$ is a unit; it follows that $x_r^d \in \mathfrak{x}' + \mathfrak{x}^{d+1}$; since $\mathfrak{x}^d \subset \mathfrak{x}' + \mathfrak{o}x_r^d$, we have $\mathfrak{x}^d \subset \mathfrak{x}' + \mathfrak{x}^{d+1}$, whence, by induction, $\mathfrak{x}^d \subset \mathfrak{x}' + \mathfrak{x}^{d+n}$ for every n; by Lemma 6, §II, p. 695, it follows that $\mathfrak{x}^d \subset \mathfrak{x}'$, whence $\mathfrak{m}^{dh} \subset \mathfrak{x}'$ for some h, which would imply $\dim \mathfrak{o} < r$.

PROPOSITION 5. *Let $\mathfrak{o}$ be a complete local ring, and let $x_1, \cdots, x_m$ be elements of $\mathfrak{o}$ which generate an ideal which is primary for the ideal $\mathfrak{m}$ of non-units in $\mathfrak{o}$. Then $\mathfrak{o}$ is a finite module over $K[[x_1, \cdots, x_m]]$.*

This follows immediately from Proposition 4, §II, p. 695.

PROPOSITION 6. *Let $\mathfrak{o}$ be a local ring, and let $\mathfrak{a}$ be an ideal in $\mathfrak{o}$. If $\mathfrak{a}$ contains an element which is not a zero divisor in $\mathfrak{o}$, we have* $\dim \mathfrak{o}/\mathfrak{a} < \dim \mathfrak{o}$.

Let $\bar{\mathfrak{o}}$ be the completion of $\mathfrak{o}$; we know that $\bar{\mathfrak{o}}/\mathfrak{a}\bar{\mathfrak{o}}$ is isomorphic with the completion of $\mathfrak{o}/\mathfrak{a}$ (Proposition 5, §II, p. 699). Moreover $\mathfrak{a}\bar{\mathfrak{o}}$ contains an element which is not a zero divisor in $\bar{\mathfrak{o}}$ (Proposition 6, §II, p. 699). It follows that it

[10] Mr. Cohen has proved that every complete local ring (in the sense of the definition of Krull) which contains a field contains a field which is a complete system of representatives for the residue classes modulo the maximal ideal.

[11] It can be proved that Lemma 1 still holds for any local ring in the sense of Krull.

is sufficient to prove our assertion in the case where $\mathfrak{o}$ is complete. Let K be a basic field of $\mathfrak{o}$, and let $\{x_1, \cdots, x_r\}$ be a system of parameters in $\mathfrak{o}$. If y is an element of $\mathfrak{a}$ which is not a zero divisor, y is integral over $K[[x_1, \cdots, x_r]]$ by Proposition 5: there exists a relationship of the form $y^m + b_1 y^{m-1} + \cdots + b_m = 0$, with $b_i \in K[[x_1, \cdots, x_r]]$ $(1 \leqq i \leqq m)$; since y is not a zero-divisor, we may assume that $b_m = F(x_1, \cdots, x_r) \neq 0$. We have $b_m \in \mathfrak{a}$, whence $F(x_1', \cdots, x_r') = 0$, if $x_1', \cdots, x_r'$ are the residue classes of $x_1, \cdots, x_r$ modulo $\mathfrak{a}$. By Proposition 4 above, it follows that $x_1', \cdots, x_r'$ do not form a system of parameters in $\mathfrak{o}/\mathfrak{a}$, whence $\dim \mathfrak{o}/\mathfrak{a} < r$ by Proposition 1.

PROPOSITION 7. *Let $x_1, \cdots, x_m$ be elements of a local ring $\mathfrak{o}$ of dimension r which generate an ideal which is primary for the ideal $\mathfrak{m}$ of non-units in $\mathfrak{o}$. Let K be a basic field of $\mathfrak{o}$; it is possible to find r linear combinations of $x_1, \cdots, x_m$ with coefficients in $\mathfrak{o}$ which form a system of parameters in $\mathfrak{o}$.*

We proceed by induction on r. The result is trivial for $r = 0$. Assume that $r > 0$ and that our result holds in local rings of dimension $r - 1$. Let $\mathfrak{x}$ be the ideal $\sum_{i=1}^{m} \mathfrak{o}x_i$, and let $\mathfrak{c}$ be the set of the elements $c \in \mathfrak{o}$ such that $c\mathfrak{x}^a = \{0\}$ holds for some a. It is clear that $\mathfrak{c}$ is an ideal; since it has a finite set of generators, we have $\mathfrak{c}\mathfrak{x}^{a_0} = \{0\}$ for some a_0; if $\mathfrak{m}^h \subset \mathfrak{x}$, we have $\mathfrak{m}^{a_0 h}\mathfrak{c} = \{0\}$, whence $\mathfrak{c}^{a_0 h+1} = \{0\}$, since $\mathfrak{c} \subset \mathfrak{m}$. We set $\mathfrak{o}' = \mathfrak{o}/\mathfrak{c}$, and we denote by $x_1', \cdots, x_m'$ the residue classes of $x_1, \cdots, x_m$ modulo $\mathfrak{c}$. If $c' \in \mathfrak{o}'$, the conditions $c'x_i' = 0$ $(1 \leqq i \leqq m)$ imply $c' = 0$. Let $\mathfrak{n}_1, \cdots, \mathfrak{n}_g$ be the prime divisors of the zero ideal in $\mathfrak{o}'$: none of these ideals can contain all the elements x_i' at the same time. Let $\mathfrak{M}$ be the vector space over K spanned by $x_1', \cdots, x_m'$; if $\mathfrak{M}_i = \mathfrak{M} \cap n_i$, we have $\mathfrak{M}_i \neq \mathfrak{M}$ $(1 \leqq i \leqq g)$; since K contains infinitely many elements, we can find a linear combination y_1' of $x_1', \cdots, x_m'$ with coefficients in K which does not belong to any one of the sets $\mathfrak{M}_i$: y_1' is not a zero-divisor in $\mathfrak{o}'$. We denote by r' the dimension of $\mathfrak{o}'$, whence $r' \leqq r$. Let $\mathfrak{o}''$ be the ring $\mathfrak{o}'/\mathfrak{o}'y_1'$: we have $\dim \mathfrak{o}'' < r'$ by Proposition 5. If $y_2', \cdots, y_s'$ are elements of $\mathfrak{o}'$ whose residue classes modulo $\mathfrak{o}'y_1'$ form a system of parameters in $\mathfrak{o}''$, it is clear that the ideal $\sum_{i=1}^{s} \mathfrak{o}'y_i'$ is primary for the ideal of non-units in $\mathfrak{o}'$, whence $s + 1 \leqq r'$. It follows that $s = r' - 1$. In virtue of our induction assumption, we may take for $y_2', \cdots, y_m'$, linear combinations of $x_1', \cdots, x_m'$ with coefficients in K; the elements $y_1', \cdots, y_{r'}'$ form a system of parameters in $\mathfrak{o}'$ which is composed of linear combinations of $x_1', \cdots, x_m'$ with coefficients in K. Let $y_1, \cdots, y_{r'}$ be linear combinations of $x_1, \cdots, x_m$ with coefficients in K belonging respectively to the residue classes $y_1', \cdots, y_{r'}'$ modulo $\mathfrak{c}$, and let $\mathfrak{m}_0$ be any prime ideal containing $y_1, \cdots, y_{r'}$. The ideal of non-units in $\mathfrak{o}'$ being $\mathfrak{m}/\mathfrak{c}$, we see that $\mathfrak{m}^k \subset \mathfrak{m}_0 + \mathfrak{c}$ for some k. It follows that $\mathfrak{m}^{k(a_0 h+1)} \subset \mathfrak{m}_0$, whence $\mathfrak{m}_0 = \mathfrak{m}$, which proves that $r' = r$ and that $y_1, \cdots, y_r$ form a system of parameters in $\mathfrak{o}$.

COROLLARY. *Let $\mathfrak{o}$ be a complete local ring, and let K be a basic field of $\mathfrak{o}$. If $x_1, \cdots, x_r$ are elements of $\mathfrak{o}$ which are analytically independent over K and such that $\mathfrak{o}$ is a finite module over $K[[x_1, \cdots, x_r]]$, these elements form a system of parameters in $\mathfrak{o}$.*

Let $\mathfrak{m}$ be any prime ideal in $\mathfrak{o}$ containing $x_1, \cdots, x_r$; the ring $\mathfrak{o}/\mathfrak{m}$ is a finite

module over $K[[x_1, \cdots, x_r]]/K[[x_1, \cdots, x_r]] \cap \mathfrak{m} = K$; since $\mathfrak{o}/\mathfrak{m}$ is a ring without zero divisors $\neq 0$, $\mathfrak{o}/\mathfrak{m}$ is a field, which proves that $\mathfrak{m}$ is the ideal of non-units in $\mathfrak{o}$. By Proposition 7, we can form s linear combinations $y_1, \cdots, y_s$ of $x_1, \cdots, x_r$ with coefficients in K which constitute a system of parameters in $\mathfrak{o}$; since $K[[x_1, \cdots, x_r]]$ is a finite module over $K[[y_1, \cdots, y_s]]$, we must obviously have $s = r$, which proves the corollary.

PROPOSITION 8. *Let $\mathfrak{o}'$ be a ring which contains a complete local ring $\mathfrak{o}$ as a sub-ring and is a finite $\mathfrak{o}$-module. If $\mathfrak{o}'$ has no zero divisor $\neq 0$, $\mathfrak{o}'$ is a complete local ring. If $\mathfrak{m}$ is the ideal of non-units in $\mathfrak{o}$, $\mathfrak{m}\mathfrak{o}'$ is primary for the ideal of non-units in $\mathfrak{o}'$.*

By Proposition 3, §II, p. 694 we know that $\mathfrak{o}'$ is a complete semi-local ring and that $\mathfrak{m}\mathfrak{o}'$ contains a power of the product of all maximal prime ideals $\mathfrak{m}'$ of $\mathfrak{o}'$. By Proposition 2, §II, p. 693, we see that, $\mathfrak{o}'$ having no zero-divisor $\neq 0$, there can exist only one prime ideal $\mathfrak{m}'$. Proposition 8 then follows immediately.

DEFINITION 3. *A local ring $\mathfrak{o}$ is said to be regular*[12] *if the ideal $\mathfrak{m}$ of non-units in $\mathfrak{o}$ can be generated by a system of parameters; such a system of parameters is then said to be regular.*

From now on, we shall denote by $\mathfrak{o}$ a regular local ring and by $\{x_1, \cdots, x_r\}$ a regular system of parameters in $\mathfrak{o}$.

It is clear that the completion of $\mathfrak{o}$ is again regular; if $\{u_1, \cdots, u_d\}$ is a set of elements of $\mathfrak{o}$ whose residue classes modulo $\mathfrak{m}$ form a linear base of $\mathfrak{o}/\mathfrak{m}$ with respect to a basic field of $\mathfrak{o}$, we have $\bar{\mathfrak{o}} = \sum_{i=1}^{d} K[[x_1, \cdots, x_r]]u_i$.[13] Every element $a \in \mathfrak{m}^k$ may be written in the form $F(x_1, \cdots, x_r)$, where F is a form of degree k with coefficients in $\mathfrak{o}$; by Lemma 1, the form F^* deduced from F by reduction of the coefficients modulo $\mathfrak{m}$ is uniquely determined when a is given; this form is null if and only if $a \in \mathfrak{m}^{k+1}$. It follows that, if $a \in \mathfrak{m}^k$, $a \notin \mathfrak{m}^{k+1}$, $b \in \mathfrak{m}^l$, $b \notin \mathfrak{m}^{l+1}$, we have $ab \notin \mathfrak{m}^{k+l+1}$. In particular, a regular local ring has no zero divisor $\neq 0$.

Let now $\mathfrak{a}$ be the ideal generated in $\mathfrak{o}$ by $x_1, \cdots, x_{r-s}$, where $s < r$. The residue classes $x^*_{r-s+1}, \cdots, x^*_r$ of $x_{r-s+1}, \cdots, x_r$ modulo $\mathfrak{a}$ generate the ideal of non units in $\mathfrak{o}/\mathfrak{a}$; conversely, if $y_1, \cdots, y_t$ are elements of $\mathfrak{o}$ whose residue classes modulo $\mathfrak{a}$ form a system of parameters in $\mathfrak{o}/\mathfrak{a}$, the ideal generated by $x_1, \cdots, x_{r-s}, y_1, \cdots, y_t$ is primary for $\mathfrak{m}$, whence $\dim \mathfrak{o}/\mathfrak{a} = t \geqq r - s$. It follows that $\mathfrak{o}/\mathfrak{a}$ is regular and that $x^*_{r-s+1}, \cdots, x^*_r$ form a regular system of parameters in $\mathfrak{o}/\mathfrak{a}$. Since $\mathfrak{o}/\mathfrak{a}$ has no zero divisor $\neq 0$, $\mathfrak{a}$ is prime.

Conversely, let $\mathfrak{p}$ be a prime ideal in $\mathfrak{o}$ such that $\mathfrak{o}/\mathfrak{p}$ is regular, and let $y_1, \cdots, y_s$ be elements of $\mathfrak{o}$ whose residue classes modulo $\mathfrak{p}$ form a regular system of parameters in $\mathfrak{o}/\mathfrak{p}$. We can express y_i in the form $\sum_{j=1}^{n} a_{ij}x_j (a_{ij} \in \mathfrak{o})$; let a^*_{ij} be the residue class of a_{ij} modulo $\mathfrak{p}$. Since the residue classes $x^*_1, \cdots, x^*_r$ of $x_1, \cdots, x_r$ clearly form a system of generators of $\mathfrak{o}/\mathfrak{p}$, we have also $x^*_j =$

---

[12] This notion corresponds to the notion of "p. Stellenring" of Krull, l.c. note 3), p. 692. We use the word "regular" because these rings occur in the theory of the regular points of an algebraic variety.

[13] Cf. the proof of Proposition 4, §II, p. 695.

$\sum_{k=1}^{s} b_{jk}^* y_k^*$; let A^*, B^* be the rectangular matrices (a_{ij}^*), (b_{jk}^*), and let $C^* = (c_{ik}^*)$ be the matrix A^*B^*. We have $y_i^* = \sum_{k=1}^{s} c_{ik}^* y_k^*$; by Lemma 1, we have $c_{ik}^* \equiv \delta_{ik}$ (mod. $\mathfrak{m}/\mathfrak{p}$). It follows that the determinant of C^* is a unit, which proves that one at least of the determinants of order s extracted from A^* is a unit; we may assume without loss of generality that this happens for the determinant of the s last columns. We may then express $x_{r-s+1}, \cdots, x_r$ as linear combinations of $y_1, \cdots, y_s, x_1, \cdots, x_{r-s}$ with coefficients in $\mathfrak{o}$, whence $\mathfrak{m} = \sum_{i=1}^{s} \mathfrak{o} y_i + \sum_{h=1}^{r-s} \mathfrak{o} x_h$. Let b_{jk} be a representative in $\mathfrak{o}$ of the residue class b_{jk}^* modulo $\mathfrak{p}$; we set $x_h' = x_h - \sum_{k=1}^{s} b_{h,k} y_k$. It is clear that $\{y_1, \cdots, y_s, x_1', \cdots, x_{r-s}'\}$ is a regular system of parameters in $\mathfrak{o}$ and that $x_h' \in \mathfrak{p}$ $(1 \leqq h \leqq r - s)$. The ideal $\mathfrak{p}' = \sum_{h=1}^{r-s} \mathfrak{o} x_h'$ is contained in $\mathfrak{p}$; we know already that it is a prime ideal and that $\dim \mathfrak{o}/\mathfrak{p}' = s = \dim \mathfrak{o}/\mathfrak{p}$. By Proposition 6, it follows that $\mathfrak{p} = \mathfrak{p}'$. We have therefore proved the following results:

PROPOSITION 9. *Let $\mathfrak{o}$ be a regular local ring. Then $\mathfrak{o}$ does not contain any zero divisor $\neq 0$. If $\{x_1, \cdots, x_r\}$ is a regular system of parameters in $\mathfrak{o}$, and if $s < r$, the ideal generated in $\mathfrak{o}$ by $x_1, \cdots, x_{r-s}$ is prime and the factor ring of $\mathfrak{o}$ by this ideal is regular of dimension s. Conversely, let $\mathfrak{p}$ be a prime ideal in $\mathfrak{o}$ such that $\mathfrak{o}/\mathfrak{p}$ is regular, and let $y_1, \cdots, y_s$ be elements of $\mathfrak{o}$ whose residue classes modulo $\mathfrak{p}$ form a regular system of parameters in $\mathfrak{o}/\mathfrak{p}$. Then we can find $r - s$ elements in $\mathfrak{o}$ which form a set of generators of $\mathfrak{p}$ and which, together with $y_1, \cdots, y_s$, form a regular system of parameters in $\mathfrak{o}$.*

§IV. Definition of the Multiplicity

LEMMA 1. *Let $\mathfrak{S}$ be a hypercomplex system over a field K, and let $\{0\} = \bigcap_{i=1}^{g} \mathfrak{q}_i$ be an irredundant representation of the zero-ideal in $\mathfrak{S}$ as an intersection of primary ideals. If $\mathfrak{u}_i$ is the prime divisor of $\mathfrak{q}_i$, we have $[\mathfrak{S}:K] = \sum_{i=1}^{g} [\mathfrak{S}/\mathfrak{u}_i:K]\cdot l_i$, where l_i is the length of $\mathfrak{q}_i$.*

It is well known that $\mathfrak{S}$ is the direct sum of g ideals $\mathfrak{S}_i = \mathfrak{S}\epsilon_i$, where each ϵ_i is an idempotent and $\mathfrak{S}_i$ is isomorphic (as a ring) with $\mathfrak{S}/\mathfrak{q}_i$; $\mathfrak{S}_i$ may be considered as a "primary" hypercomplex system over $K\epsilon_i$, in which the only prime ideal is $\mathfrak{u}_i/\mathfrak{q}_i$. It follows that it will be sufficient to prove the lemma in the case where $g = 1$. We then set $\mathfrak{u}_1 = \mathfrak{u}$, $l_1 = l$, and we have to prove that $[\mathfrak{S}:K] = [\mathfrak{S}/\mathfrak{u}:R]\cdot l$. Let us form a chain of $l + 1$ ideals $\mathfrak{v}_0, \mathfrak{v}_1, \cdots, \mathfrak{v}_l$, beginning with $\mathfrak{v}_0 = \{0\}$, ending with $\mathfrak{v}_l = \mathfrak{S}$, such that $\mathfrak{v}_{j-1} \subset \mathfrak{v}_j$, $\mathfrak{v}_{j-1} \neq \mathfrak{v}_j$ and that no further ideal can be inserted between $\mathfrak{v}_{j-1}$ and $\mathfrak{v}_j$ $(1 \leqq j \leqq l)$. Let v_j $(1 \leqq j \leqq l)$ be an element of $\mathfrak{v}_j$ not contained in $\mathfrak{v}_{j-1}$; we have clearly $\mathfrak{v}_j = \mathfrak{S}v_1 + \cdots + \mathfrak{S}v_j$. We shall prove that $\mathfrak{u}\mathfrak{v}_j \subset \mathfrak{v}_{j-1}$; in fact, the ideal $\mathfrak{v}_{j-1} + \mathfrak{u}\mathfrak{v}_j$ is either equal to $\mathfrak{v}_{j-1}$ (in which case our assertion is proved) or to $\mathfrak{v}_j$; in the latter case, we would have $v_j(1 + u) \in \mathfrak{v}_{j-1}$ for some $u \in \mathfrak{u}$, which is impossible because $1 + u$ is clearly a unit in $\mathfrak{S}$. It follows that $\mathfrak{v}_j\mathfrak{v}/_{j-1}$ may be considered as an $(\mathfrak{S}/\mathfrak{u})$-module. Every $(\mathfrak{S}/\mathfrak{u})$-sub-module of $\mathfrak{v}_j/\mathfrak{v}_{j-1}$ is of the form $\mathfrak{v}/\mathfrak{v}_{j-1}$ where $\mathfrak{v}$ is an ideal in $\mathfrak{S}$ such that $\mathfrak{v}_{j-1} \subset \mathfrak{v} \subset \mathfrak{v}_j$; therefore $\mathfrak{v}/_j\mathfrak{p}_{j-1}$ is an $(\mathfrak{S}/\mathfrak{u})$-module of dimension 1, whence $[\mathfrak{v}_j/\mathfrak{v}_{j-1}:K] = [\mathfrak{S}/\mathfrak{u}:K]$. Since $[\mathfrak{S}:K] = \sum_{j=1}^{g} [\mathfrak{v}_j/\mathfrak{v}_{j-1}:K]$, Lemma 1 is proved.

LEMMA 2. *Let $\mathfrak{r}$ be a local ring of dimension 1 in which the ideal of non units*

is principal, and let x be a generator of this ideal. Let $\mathfrak{o}$ be a ring containing $\mathfrak{r}$ and which is a finite $\mathfrak{r}$-module. Assume that no element $\neq 0$ if $\mathfrak{r}$ is a zero divisor in $\mathfrak{o}$. Let $x\mathfrak{o} = \bigcap_{i=1}^{g} \mathfrak{q}_i$ be an irredundant representation of $x\mathfrak{o}$ as an intersection of primary ideals in $\mathfrak{o}$. We have $[\mathfrak{o}:\mathfrak{r}] = \sum_{i=1}^{g} [\mathfrak{o}/\mathfrak{u}_i:\mathfrak{r}/x\mathfrak{r}]\cdot l_i$, where $\mathfrak{u}_i$ is the associated prime ideal of $\mathfrak{q}_i$ and where l_i is the length of $\mathfrak{q}_i$.[14]

It is clear that every ideal in $\mathfrak{r}$ is principal and is generated by some power of x. It follows that $\mathfrak{o}$ has a linear base $\{\omega_1, \cdots, \omega_d\}$ over $\mathfrak{r}$ (the elements $\omega_1, \cdots, \omega_d$ being linearly independent over $\mathfrak{r}$). Since $\mathfrak{u}_i \neq \mathfrak{o}$ and since $x\mathfrak{r}$ is a maximal prime ideal in $\mathfrak{r}$, we have $\mathfrak{u}_i \cap \mathfrak{r} = x\mathfrak{r}$ $(1 \leqq i \leqq g)$. The ring $\mathfrak{o}/x\mathfrak{o}$ is a finite module over the field $\mathfrak{r}/x\mathfrak{r}$; this means that $\mathfrak{o}/x\mathfrak{o}$ is a hypercomplex system over $\mathfrak{r}/x\mathfrak{r}$. It is clear that the residue classes of $\omega_1, \cdots, \omega_d$ modulo $x\mathfrak{o}$ form a linear base of $\mathfrak{o}/x\mathfrak{o}$ over $\mathfrak{r}/x\mathfrak{r}$, whence $[\mathfrak{o}/x\mathfrak{o}:\mathfrak{r}/x\mathfrak{r}] = [\mathfrak{o}:\mathfrak{r}]$.

It is clear that $\{0\} = \bigcap_{i=1}^{g} \mathfrak{q}_i/x\mathfrak{o}$ is an irredundant representation of the zero ideal in $\mathfrak{o}/x\mathfrak{o}$ as an intersection of primary ideals and that the associated prime ideal of $\mathfrak{q}_i/x\mathfrak{o}$ is $\mathfrak{u}_i/x\mathfrak{o}$. The length of $\mathfrak{q}_i/x\mathfrak{o}$ is equal to l_i, because there exists a one-to-one inclusion preserving correspondence between the ideals containing $\mathfrak{q}_i$ in $\mathfrak{o}$ and the ideals containing $\mathfrak{q}_i/x\mathfrak{o}$ in $\mathfrak{o}/x\mathfrak{o}$. The field $(\mathfrak{o}/x\mathfrak{o})/(\mathfrak{u}_i/x\mathfrak{o})$ is isomorphic to $\mathfrak{o}/\mathfrak{u}_i$; therefore Lemma 2 follows immediately from Lemma 1.

THEOREM 2. *Let $\mathfrak{o}$ be a complete local ring which does not contain any zero divisor $\neq 0$, and let $\{x_1, \cdots, x_r\}$ be a system of parameters in $\mathfrak{o}$. There exists an integer $e = e(\mathfrak{o}; x_1, \cdots, x_r)$ which has the following property; if K is any basic field of $\mathfrak{o}$, we have $[\mathfrak{o}:K[[x_1, \cdots, x_r]]] = e.\ [\mathfrak{o}/\mathfrak{m}:K]$, where $\mathfrak{m}$ is the ideal of non units in $\mathfrak{o}$.*

The proof proceeds by induction on the dimension r of $\mathfrak{o}$. If $r = 1$, the result follows immediately from Lemma 2. Assume that $r > 1$ and that the result holds for the dimension $r - 1$. If K is a basic field of $\mathfrak{o}$, an element of the form $x_r^{-1}F(x_1, \cdots, x_{r-1})$ (where F is a power series with coefficients in K) is not integral over $K[[x_1, \cdots, x_r]]$ (because this ring is integrally closed) and therefore does not belong to $\mathfrak{o}$; it follows that $x_r\mathfrak{o} \cap K[[x_1, \cdots, x_r]] = x_rK[[x_1, \cdots, x_r]]$. The ideal $x_r\mathfrak{o}$ contains some power product of prime ideals in $\mathfrak{o}$; if every one of these prime ideals were to contain an element $\neq 0$ of $K[[x_1, \cdots, x_{r-1}]]$ the same would hold for $x_r\mathfrak{o}$ (because $K[[x_1, \cdots, x_{r-1}]]$ has no zero-divisor $\neq 0$); this is not being the case, there exists at least one prime ideal $\mathfrak{u}$ in $\mathfrak{o}$ containing x_r and such that $\mathfrak{u} \cap K[[x_1, \cdots, x_{r-1}]] = \{0\}$. If $\mathfrak{u}$ is any such ideal, the residue classes $x_1', \cdots, x_{r-1}'$ of $x_1, \cdots, x_{r-1}$ modulo $\mathfrak{u}$ are analytically independent over K and $\mathfrak{o}/\mathfrak{u}$ is a finite module over the ring $K[[x_1', \cdots, x_{r-1}']]$, whence $\dim \mathfrak{o}/\mathfrak{u} = r - 1$, by the corollary to Proposition 7, §III, p. 703. Conversely, if $\mathfrak{u}$ is a prime ideal containing x_r and such that $\dim \mathfrak{o}/\mathfrak{u} = r - 1$, the residue classes of $x_1, \cdots, x_{r-1}$ modulo $\mathfrak{u}$ obviously form a system of parameters in $\mathfrak{o}/\mathfrak{u}$, whence $\mathfrak{u} \cap K[[x_1, \cdots, x_{r-1}]] = \{0\}$. We see that the prime ideals $\mathfrak{u}$ containing x_r and which satisfy the latter condition are characterized independently of the basic field K; it follows immediately from Proposition 6, §III, p. 702 that any

[14] We denote by $[\mathfrak{o}:\mathfrak{r}]$ the maximum number of elements of $\mathfrak{o}$ linearly independent over $\mathfrak{r}$. This is also the dimension of the ring of quotients of $\mathfrak{o}$, considered as a vector space over the field of quotients of $\mathfrak{r}$.

one of these ideals is a minimal prime divisor of $x_r\mathfrak{o}$, which proves that there are only a finite number of them, say $\mathfrak{u}_1, \cdots, \mathfrak{u}_g$; moreover there corresponds to every $\mathfrak{u}_j$ a uniquely defined primary component $\mathfrak{v}_j$ of $x_r\mathfrak{o}$; we denote by l_j the length of $\mathfrak{v}_j$. Since $\mathfrak{o}/\mathfrak{u}_j$ is of dimension $r - 1$, our induction assumption guarantees the existence of a number $e_j = e(\mathfrak{o}/\mathfrak{u}_j; x_{1,j}, \cdots, x_{r-1,j})$ (where $x_{i,j}$ is the residue class of x_i modulo $\mathfrak{u}_j$) such that $[\mathfrak{o}/\mathfrak{u}_j:K[[x_{1,j}, \cdots, x_{r-1,j}]]] = e_j \cdot [(\mathfrak{o}/\mathfrak{u}_j)/(\mathfrak{m}/\mathfrak{u}_j):K]$. We shall see that the number $e = \sum_j e_j l_j$ has the required property.

We denote by S the set of the elements $\neq 0$ in $K[[x_1, \cdots, x_{r-1}]]$; S is multiplicatively closed and no element of S is a zero divisor in $\mathfrak{o}$; it follows that we can construct the rings of quotients $\mathfrak{o}_S$ of $\mathfrak{o}$ and $\mathfrak{r}$ of $K[[x_1, \cdots, x_r]]$ with respect to S. The ideals $\mathfrak{u}_j$ are all the prime divisors of $x\mathfrak{o}$ which do not meet S; by Lemma 3, §I, p. 691, $x_r\mathfrak{o}_S = \bigcap_{j=1}^{g} \mathfrak{v}_j\mathfrak{o}_S$ is an irredundant representation of $x_r\mathfrak{o}$ as intersection of primary ideals in $\mathfrak{o}_S$, and $\mathfrak{v}_j\mathfrak{o}_S$ is of length l_j. The ring $\mathfrak{r}$ is clearly a local ring of dimension 1 in which the ideal of non units is generated by x_r; since $\mathfrak{o}$ is finite over $K[[x_1, \cdots, x_r]]$, $\mathfrak{o}_S$ is finite over $\mathfrak{r}$. The field $\mathfrak{o}_S/\mathfrak{u}_j\mathfrak{o}_S$ is the field of quotients of $\mathfrak{o}/\mathfrak{u}_j$, whence $[\mathfrak{o}_S/\mathfrak{u}_j\mathfrak{o}_S:\mathfrak{r}/\mathfrak{u}_j\mathfrak{o}_S \cap \mathfrak{r}] = e_j \cdot [(\mathfrak{o}/\mathfrak{u}_j)/(\mathfrak{m}/\mathfrak{u}_j):K] = e_j[\mathfrak{o}/\mathfrak{m}:K]$. By Lemma 2, we have $[\mathfrak{o}_S:\mathfrak{r}] = e[\mathfrak{o}/\mathfrak{m}:K]$, which proves Theorem 2 for the dimension r, since $[\mathfrak{o}_S:\mathfrak{r}] = [\mathfrak{o}:K[[x_1, \cdots, x_r]]]$.

Definition 1. *If $\mathfrak{o}$ is a complete local ring without any zero divisor $\neq 0$, and if $\{x_1, \cdots, x_r\}$ is a system of parameters in $\mathfrak{o}$, the number $e(\mathfrak{o}; x_1, \cdots, x_r)$ whose existence is asserted in Theorem 2 is called the multiplicity of $\mathfrak{o}$ with respect to the system $\{x_1, \cdots, x_r\}$.*

It is clear that this multiplicity is an invariant of the ring $K[[x_1, \cdots, x_r]]$, where K is any basic field of $\mathfrak{o}$. We shall prove that it is also an invariant of the ideal $\sum_{i=1}^r \mathfrak{o}x_i$.

Proposition 1. *Let $\mathfrak{o}$ be a complete local ring which does not contain any zero divisor $\neq 0$. If $\{x_1, \cdots, x_r\}$ and $\{x'_1, \cdots, x'_r\}$ are two systems of parameters in $\mathfrak{o}$, the equality $\sum_1^r \mathfrak{o}x_i = \sum_1^r \mathfrak{o}x'_i$ implies $e(\mathfrak{o}; x_1, \cdots, x_r) = e(\mathfrak{o}; x'_1, \cdots, x'_r)$.*

We set $x'_i = \sum_{j=1}^r a_{ij}x_j$ ($a_{ij} \in \mathfrak{o}$), and we denote by $\mathfrak{m}$ the ideal of non-units in $\mathfrak{o}$. One at least of the elements a_{ij} does not belong to $\mathfrak{m}$, because otherwise we would have $\sum_1^r \mathfrak{o}x'_i \subset \mathfrak{m}(\sum_1^r \mathfrak{o}x_i)$, which is clearly impossible if $\sum_1^r \mathfrak{o}x_i = \sum_1^r \mathfrak{o}x'_i$. We may assume without loss of generality that $a_{rr} \notin \mathfrak{m}$; it follows that a_{rr} is a unit, whence $x_r \in \sum_{i=1}^{r-1} x_i\mathfrak{o} + x'_r\mathfrak{o}$ and $\sum_1^r \mathfrak{o}x_i = \sum_{i=1}^{r-1} \mathfrak{o}x_i + \mathfrak{o}x'_r$. It follows that, if $r > 1$, it is sufficient to prove Proposition 1 under the supplementary assumption that $\{x_1, \cdots, x_r\}$ and $\{x'_1, \cdots, x'_r\}$ have an element in common.

This being said, we prove Proposition 1 by induction on r. If $r = 1$, Lemma 2 shows that $e(\mathfrak{o}; x_1)$ is the length of the primary ideal $\mathfrak{o}x_1$, which proves Proposition 1 in this case. Assume now that $r > 1$ and that $x'_r = x_r$. Making use of the notations of the proof of Theorem 2, we see that it is sufficient to prove that $e(\mathfrak{o}/\mathfrak{u}_j; x_{1,j}, \cdots, x_{r-1,j}) = e(\mathfrak{o}/\mathfrak{u}_j; x'_{1,j}, \cdots, x'_{r-1,j})$, where $x_{i,j}$, $x'_{i,j}$ are the residue classes of x_i, x'_i modulo $\mathfrak{u}_j$. But we have $\sum_{i=1}^{r-1} (\mathfrak{o}/\mathfrak{u}_j)x_{i,j} = (\mathfrak{u}_j + \sum_{i=1}^r \mathfrak{o}x_i)/\mathfrak{u}_j = (\mathfrak{u}_j + \sum_{i=1}^r \mathfrak{o}x'_i)/\mathfrak{u}_j = \sum_{i=1}^{r-1} (\mathfrak{o}/\mathfrak{u}_j)x'_{i,j}$, which proves the formula in question, since $\dim \mathfrak{o}/\mathfrak{u}_j = r - 1$.

APPENDIX

On the dimension of local rings

We shall prove that the definition of the dimension of a local ring which is given in the text (Definition 2, §III, p. 701) is equivalent to the definition given by Krull (loc. cit., note 3, p. 692). In this appendix, we shall take the notion of local ring in the sense of Krull (i.e., we drop from our definition the requirement of existence of a basic field).

Let $\mathfrak{o}$ be a local ring; we denote by r the dimension of $\mathfrak{o}$ in the sense of our definition and by r' its dimension in the sense of Krull's definition. The inequality $r' \leqq r$ follows immediately from Theorem 7* in Krull's paper; we have therefore to prove that $r \leqq r'$; it will be sufficient to prove that there exists a chain of $r + 1$ prime ideals $\mathfrak{p}_0, \mathfrak{p}_1, \cdots, \mathfrak{p}_r$ in $\mathfrak{o}$ such that $\mathfrak{p}_{i-1}$ is properly contained in $\mathfrak{p}_i$ $(1 \leqq i \leqq r)$. There is nothing to prove if $r = 0$. Assume that $r > 0$ and that our assertion holds for local rings of dimension $r - 1$. We denote by $\mathfrak{p}$ the ideal of non units in $\mathfrak{o}$ and by $\mathfrak{n}_1, \cdots, \mathfrak{n}_g$ the distinct minimal prime divisors of the zero ideal in $\mathfrak{o}$. We have $\mathfrak{n}_i \neq \mathfrak{p}$ $(1 \leqq i \leqq g)$ because $r \neq 0$; it follows that none of the ideals $\mathfrak{n}_i$ can contain simultaneously all elements of a system of parameters $\{x_1, \cdots, x_r\}$ in $\mathfrak{o}$. Assume for instance that $x_1 \notin \mathfrak{n}_1$. It may happen that x_1 belongs to some $\mathfrak{n}_k$ with $k > 1$; if this is the case, we select an index $i(k)$ such that $x_{i(k)} \notin \mathfrak{n}_k$ and an element a_k which belongs to $\prod_{j \neq k} \mathfrak{n}_j$ but not to $\mathfrak{n}_k$ (this is possible because $\mathfrak{n}_k$ does not contain $\mathfrak{n}_j$ for $j \neq k$); we set $x_1' = x_1 + \sum_k a_k x_{i(k)}$, the sum being extended to all indices k such that $x_1 \in \mathfrak{n}_k$. The element x_1' does not belong to any of the ideals $\mathfrak{n}_i$ $(1 \leqq i \leqq g)$ and the ideal $\mathfrak{o}x_1' + \sum_{i=2}^r \mathfrak{o}x_i$ is equal to $\sum_1^r \mathfrak{o}x_i$, which proves that $x_1', x_2, \cdots, x_r$ form a system of parameters in $\mathfrak{o}$. Let $\mathfrak{o}^*$ be the local ring $\mathfrak{o}/\mathfrak{o}x_1'$; if $y_1, \cdots, y_s$ are elements of $\mathfrak{o}$ whose residue classes modulo $\mathfrak{o}x_1'$ form a system of parameters in $\mathfrak{o}^*$, there exists an exponent h such that $\mathfrak{p}^h + \mathfrak{o}x_1' \subset \mathfrak{o}x_1' + \sum_{j=1}^s \mathfrak{o}y_j$, and therefore the ideal $\mathfrak{o}x_1' + \sum_{j=1}^s \mathfrak{o}y_j$ is primary for $\mathfrak{p}$, whence $s \geqq r - 1$. Since the residue classes of $x_2, \cdots, x_r$ clearly generate an ideal in $\mathfrak{o}^*$ which is primary for $\mathfrak{p} + \mathfrak{o}x_1'/\mathfrak{o}x_1'$, we conclude that $\mathfrak{o}^*$ is of dimension $r - 1$. It follows from our inductive assumption (applied to $\mathfrak{o}^*$) that there exists a chain of r distinct prime ideals $\mathfrak{p}_1, \cdots, \mathfrak{p}_r$ in $\mathfrak{o}$, containing $\mathfrak{o}x_1'$ and such that $\mathfrak{p}_{i-1} \subset \mathfrak{p}_i$ $(2 \leqq i \leqq r)$. Since $\mathfrak{p}_1$ contains x_1', it cannot be a minimal prime divisor of the zero ideal. Since $\mathfrak{p}_1$ is prime and contains 0, it contains some minimal prime divisor of the zero ideal, say $\mathfrak{p}_0$, and our assertion is proved for local rings of dimension r.

PRINCETON UNIVERSITY

ANNALS OF MATHEMATIC
Vol. 44, No. 4, October, 1943

ON SOME ALGEBRAICAL PROPERTIES OF OPERATOR RINGS

BY JOHN VON NEUMANN

(Received February 24, 1943)

§1. The notations to be used in this paper agree with those of the papers quoted below, especially [2]. The results which we obtain will be used in [5], but they seem to have a certain interest of their own as well.

BIBLIOGRAPHY

[1] J. VON NEUMANN: "*Zur Algebra der Funktionaloperatoren . . .*," Math. Annalen, vol. 102, pp. 370–427 (1929).
[2] J. VON NEUMANN: "*Über adjungierte Funktionaloperatoren,*" Annals of Math., vol. 33, pp. 294–310 (1932).
[3] F. J. MURRAY: "*Linear transformations between Hilbert spaces,*" Trans. Amer. Math. Soc., vol. 37, 301–338 (1935).
[4] F. J. MURRAY AND J. VON NEUMANN: "*On rings of operators,*" Annals of Math., vol. 37, pp. 116–229 (1936).
[5] F. J. MURRAY AND J. VON NEUMANN: "*On rings of operators IV,*" Annals of Math., vol. 44, (following immediately).

§2. We consider an operator ring $\mathbf{M}$ in a Hilbert space $\mathfrak{H}$ which contains 1. We say that a notion defined in $\mathbf{M}$ is *purely algebraical* if it can be expressed in terms of the entity 1 (the unit operator) and the operations αA (α any complex number), A^*, $A + B$, AB alone and referring only to operators belonging to $\mathbf{M}$. Thus if a mapping of $\mathbf{M}$ onto another ring is isomorphic, that is if the notions of 1, αA, A^*, $A + B$, AB are invariant, then every "purely algebraic" notion is invariant.

Notions which refer explicitly to elements of $\mathfrak{H}$ will not in general be purely algebraical.

Now the notions which follow do refer to elements of $\mathfrak{H}$.

(α) Definiteness of an $A \in \mathbf{M}$.

(β) The numerical value of the bound $|||\, A \,|||$ of an $A \in \mathbf{M}$.

(γ) The fact that $\lim_{n\to\infty}$ strong $A_n = A$ when $A, A_1, A_2, \cdots \in \mathbf{M}$.

(δ) The fact that $\lim_{n\to\infty}$ weak $A_n = A$ when $A, A_1, A_2, \cdots \in \mathbf{M}$.

(For the *strong* and *weak* notions of convergence, used above, and for the corresponding topologies, to be referred to below, cf. [1], pp. 381–388.) The object of this paper is to show that the notions (α)–(δ) are nevertheless purely algebraical.

We shall not prove the same thing concerning the strong and weak topology (for operators). It is probably not generally true. For an important special case where it is true, cf. [5], Theorem I.

(α) and (β) are easy to dispose of, cf. §3. (δ) follows from (γ) by a known argument which will be given in §6. So the main difficulty lies in establishing the character of (γ), which will be done in §5. The discussion of (γ) and (δ)

would be much easier if the notion of being purely algebraic did not exclude the use of operators not in **M**, cf. §4.

§3. We begin by considering (α) and (β).

LEMMA 1. *$A \epsilon \mathbf{M}$ is definite if and only if $A = B^*B$ for some $B \epsilon \mathbf{M}$.*

PROOF. Sufficiency: Suppose $A = B^*B$ for $B \epsilon \mathbf{M}$. Since $B^* \epsilon \mathbf{M}$, $A \epsilon \mathbf{M}$. Furthermore $(Af, f) = (B^*Bf, f) = (Bf, Bf) = \| Bf \|^2 \geqq 0$ and hence A is definite.

Necessity: If $A \epsilon \mathbf{M}$ is definite, then there exists a definite $B \epsilon \mathbf{M}$ with $B^2 = A$. (For the existence of B cf. [2], p. 307, Theorem 7, or [4], p. 142, §4.4. B is bounded along with A.[1] That $A \epsilon \mathbf{M}$ implies $B \epsilon \mathbf{M}$ is shown in [4], p. 143, Lemma 4.4.1. There exist also simple direct proofs, using functions of functional operators.) Now $B = B^*$ so $A = B^*B$.

LEMMA 2. *$||| A |||$ is the smallest (real) number $\alpha \geqq 0$ such that $\alpha^2 \cdot 1 - A^*A$ is definite.*

$||| A |||$ is the smallest $\alpha \geqq 0$ such that for every f, $\| Af \|^2 \leqq \alpha^2 \cdot \| f \|^2$. The inequality may be rewritten $0 \leqq \alpha^2 \cdot \| f \|^2 - \| Af \|^2 = \alpha^2(f, f) - (Af, Af) = \alpha^2(f, f) - (A^*Af, f) = (\{\alpha^2 \cdot 1 - A^*A\}f, f)$. Thus the statement, "for every f, $\| Af \|^2 \leqq \alpha^2 \cdot \| f \|^2$" is equivalent to "$\alpha^2 \cdot 1 - A^*A$ is definite." Substituting in the first sentence yields the lemma.

Thus we have shown

THEOREM I. *The notions (α) and (β) (definiteness and bound) are purely algebraical.*

§4. Let us now interrupt our discussion in order to analyze (γ) and (δ) without the necessity of avoiding the use of operators not in **M**. This is much easier than the discussion with the original observance, to be given in the two next sections.

$\lim_{n\to\infty}$ strong $A_n = 0$ means $\lim_{n\to\infty} \| A_n f \| = 0$ for all $f \epsilon \mathfrak{H}$; $\lim_{n\to\infty}$ weak $A_n = 0$ means $\lim_{n\to\infty} | (A_n f, g) | = 0$ for every f and g in $\mathfrak{H}$. Clearly we may restrict ourselves in both cases to the f and g with $\| f \| = \| g \| = 1$.

Denote the closed linear set of all αf as usual by $[f]$, and its projection by $P_{[f]}$. Then one sees immediately that (for $\| f \| = \| g \| = 1$)

$$AP_{[f]}h = (f, h)Af, \qquad P_{[g]}AP_{[f]}h = (f, h) \cdot (Af, g)g.$$

Hence

$$||| AP_{[f]} ||| = \| Af \|, \qquad ||| P_{[g]}AP_{[f]} ||| = | (Af, g) | .$$

Now the $P_{[f]}$ are obviously the minimal projections of the ring **B** of all bounded operators in $\mathfrak{H}$, i.e. those projections E for which the only projections $F \leqq E$

---

[1] We have $\| Bf \|^2 = (B^*Bf, f) = (Af, f) \leqq \| Af \| \cdot \| f \| \leqq ||| A ||| \cdot \| f \|^2$. Thus $\| Bf \| \leqq \sqrt{||| A |||} \cdot \| f \|$. Hence B is bounded along with A.

are $F = 0, E$. (Cf. [4], pp. 143, 144, Definition 5.1.2. The assertion concerning minimal projections is obvious.) So we see

$$\lim_{n\to\infty} \text{strong } A_n = 0 \quad \text{means that always} \quad \lim_{n\to\infty} |||\, A_n E \,||| = 0$$

$$\lim_{n\to\infty} \text{weak } A_n = 0 \quad \text{means that always} \quad \lim_{n\to\infty} |||\, G A_n E \,||| = 0.$$

Here E, G run over all minimal projections of the ring **B** of all bounded operators in $\mathfrak{H}$.

The drawback in all this is that we had to refer to the ring **B** instead of **M**.

§5. We now proceed to the more difficult analysis of (γ) in the original sense.

DEFINITION. *A sequence $A_1, A_2, \cdots \in \mathbf{M}$ is a Σ sequence if it possesses the properties*

(i) *The (numerical) sequence $|||\, A_1 \,|||, |||\, A_2 \,|||, \cdots$ is bounded.*

(ii) *There exists an operator $X \in \mathbf{M}$ such that* 1.) *for all $C \in \mathbf{M}$ $CX = 0$ implies $C = 0$, and* 2.) *all the operators,*

$$1 - \sum_{m=1}^{n} (A_m X)^*(A_m X), \qquad (n = 1, 2, \cdots) \quad \textit{are definite.}$$

LEMMA 3. *For every Σ sequence, $A_1, A_2, \cdots \in \mathbf{M}$ we have* $\lim_{n\to\infty} \text{strong } A_n = 0$.

PROOF. For every $f \in \mathfrak{H}$ we have

$$(\{1 - \sum_{m=1}^{n} (A_m X)^*(A_m X)\}f, f) \geqq 0,$$

i.e.

$$\sum_{m=1}^{n} ((A_m X)^*(A_m X)f, f) \leqq (f, f),$$

i.e.

$$\sum_{m=1}^{n} \|\, A_m X f \,\|^2 \leqq \|\, f \,\|^2.$$

Consequently $\sum_{m=1}^{\infty} \|\, A_m X f \,\|^2 \leqq \|\, f \,\|^2$ and therefore $\lim_{m\to\infty} \|\, A_m X f \,\| = 0$.

Thus we have shown: The set $\mathfrak{S}$ of all $g \in \mathfrak{H}$ with $\lim_{m\to\infty} \|\, A_m g \,\| = 0$ contains the range of X.

$\mathfrak{S}$ is clearly a linear set, and since the $A_1, A_2, \cdots$ are uniformly bounded (by (i)), $\mathfrak{S}$ is closed. Let E be the projection on $\mathfrak{S}$. We next show $\mathfrak{S}\ \eta\ \mathbf{M}$. (For this notation, cf. [4], p. 141, Def. 4.2.1.) For suppose $U' \in \mathbf{M}'$ is unitary and $f \in \mathfrak{S}$. Then $\lim_{m\to\infty} \|\, A_m U' f \,\| = \lim_{m\to\infty} \|\, U' A_m f \,\| = \lim_{m\to\infty} \|\, A_m f \,\| = 0$. Thus $f \in \mathfrak{S}$ implies $U'f \in \mathfrak{S}$ and $\mathfrak{S}$ is invariant under every $U' \in \mathbf{M}'$ or $\mathfrak{S}\ \eta\ \mathbf{M}$. This implies $E \in \mathbf{M}$.

Since the range of X is contained in $\mathfrak{S}$, $EX = X$ or $(1 - E)X = 0$. Hence by (ii), $1 - E = 0$ and $1 = E$. Thus $\mathfrak{S} = \mathfrak{H}$ that is $\lim_{m\to\infty} \text{strong } A_m = 0$.

LEMMA 4. *For every sequence, $A_1, A_2, \cdots \in \mathbf{M}$ with* $\lim_{m\to\infty} \text{strong } A_n = 0$ *there exists a subsequence $A_{1'}, A_{2'}, \cdots$ which is a Σ sequence.*

PROOF. $|||\, A_1 \,|||, |||\, A_2 \,|||, \cdots$ is bounded by [1], p. 382, footnote 35), hence $|||\, A_{1'} \,|||, |||\, A_{2'} \,|||, \cdots$ is a *fortiori* bounded for every subsequence.

Consider now an everywhere dense sequence $f_1^0, f_2^0, \cdots$ in $\mathfrak{H}$.

For every $i = 1, 2, \cdots, \lim_{n\to\infty} \|\, A_n f_i^0 \,\| = 0$. Choose accordingly $k(i)$ such

that for $n \geqq k(i)$, $\| A_n f_j^0 \| \leqq 1/2^i$ for all $j = 1, \cdots, i$. Choose a subsequence $1', 2', \cdots$ of $1, 2, \cdots$ with $1' < 2' < \cdots$ and $i' \geqq k(i)$. Then $\| A_{n'} f_j^0 \| \leqq 1/2^n$ if $n \geqq j$. Consequently $\sum_{n=1}^{\infty} \| A_{n'} f_j^0 \|^2$ is finite for $j = 1, 2, \cdots$.

Thus: The set $\mathfrak{F}$ of all $g \epsilon \mathfrak{H}$ for which $\sum_{n=1}^{\infty} \| A_{n'} g \|^2$ is finite, contains all $f_1^0, f_2^0, \cdots$ and therefore it is everywhere dense in $\mathfrak{H}$.

Form the space $\infty \otimes \mathfrak{H}$ of all sequences $\langle f_1, f_2, \cdots \rangle$ of elements of $\mathfrak{H}$ with $\sum_{i=1}^{\infty} \| f_i \|^2 < \infty$. $\infty \otimes \mathfrak{H}$ is a Hilbert space. (Cf. [4], pp. 136–137, §2.4.) Consider the following operator T from $\mathfrak{H}$ to $\infty \otimes \mathfrak{H}$.

$$Tf = \langle f, A_{1'} f, A_{2'} f, \cdots \rangle$$

(Cf. [3], Def. 1.2, p. 303.) The domain of this T is obviously the above defined set $\mathfrak{F}$. T is linear and it is also easy to verify that T is closed.[2] Therefore we may apply an argument for operators between Hilbert spaces similar to that of [2], p. 309, Satz 10, according to which there exists one and only one self-adjoint (i.e. hypermaximal) definite operator B in $\mathfrak{H}$ which is metrically equivalent to T, i.e. such that it has the same domain as T and always $\| Bf \| = \| Tf \|$. (Cf. [3], §4, pp. 311, 312.)

Thus the domain of B is $\mathfrak{F}$ and always

$$\| Bf \|^2 = \| f \|^2 + \sum_{i=1}^{\infty} \| A_{i'} f \|^2. \tag{1}$$

Consider a unitary $U' \epsilon \mathbf{M}'$. Then $U' A_{i'} = A_{i'} U'$ and hence (1) implies $\| BU'f \| = \| Bf \|$ and $\| U'^{-1} BU' f \| = \| Bf \|$. Thus $U'^{-1}BU'$ possesses the above properties which characterize B uniquely. So $U'^{-1}BU' = B$, i.e. $U'B = BU'$. Thus $B \, \eta \, \mathbf{M}$. (Cf. again [4], p. 141, Def. 4.2.1.)

By (1), $Bf = 0$ implies $f = 0$. This and the self-adjointness of B imply that B^{-1} is also self-adjoint. By (1), $\| Bf \| \geqq \| f \|$, hence $\| B^{-1} g \| \leqq \| g \|$ and so B^{-1} is bounded. Now $B \, \eta \, \mathbf{M}$ yields $B^{-1} \, \eta \, \mathbf{M}$ and since B^{-1} is bounded, $B^{-1} \epsilon \mathbf{M}$. (Cf. [4], p. 141, Lemma 4.2.1.) Put $X = B^{-1}$. Thus $X \epsilon \mathbf{M}$. Since $X = B^{-1}$, $CX = 0$ implies $Cf = CXBf = (CX) \cdot Bf = 0$ for f in the domain of B which is dense. For C bounded, this implies $C = 0$ and thus we have (ii) 1 of the definition of this section.

---

[2] We must prove $\lim_{n \to \infty} \| f_n - f \| = 0$, $\lim_{n \to \infty} \| Tf_n - f^* \| = 0$ imply the existence of Tf and $Tf = f^*$.

Put $f^* = \langle f^{(0)}, f^{(1)}, f^{(2)}, \cdots \rangle$. Then $\lim_{n \to \infty} \| Tf_n - f^* \|^2 = 0$ means that

$$\lim_{n \to \infty} (\| f_n - f^{(0)} \|^2 + \sum_{i=1}^{\infty} \| A_{i'} f_n - f^{(i)} \|^2) = 0.$$

Hence, *a fortiori*

$$\lim_{n \to \infty} \| f_n - f^{(0)} \| = 0, \qquad \lim_{n \to \infty} \| A_{i'} f_n - f^{(i)} \| = 0.$$

Thus $f = f^{(0)}$ and $A_{i'} f = f^{(i)}$. Consequently

$$\| f \|^2 + \sum_{i=1}^{\infty} \| A_{i'} f \|^2 = \| f^0 \|^2 + \sum_{i=1}^{\infty} \| f^{(i)} \|^2 = \| f^* \|^2$$

i.e. $\sum_{i=1}^{\infty} \| A_{i'} f \|^2$ is finite.

So Tf is defined and $Tf = \langle f, A_{1'} f, A_{2'} f, \cdots \rangle = \langle f^{(0)}, f^{(1)}, f^{(2)}, \cdots \rangle = f^*$ as desired.

By equation (1), $\sum_{m=1}^{n} \| A_m f \|^2 \leqq \| Bf \|^2$, hence $\sum_{m=1}^{n} \| A_{m'} X g \|^2 \leqq \| g \|^2$. This may be written

$$\sum_{m=1}^{n} ((A_{m'} X)^*(A_{m'} X)g, g) \leqq (g, g)$$

or

$$(\{1 - \sum_{m=1}^{n} (A_{m'} X)^*(A_{m'} X)\}g, g) \geqq 0.$$

Thus the operator $1 - \sum_{m=1}^{n} (A_{m'}X)^*(A_{m'}X)$ is definite, i.e. we have (ii)2. This completes the proof.

LEMMA 5. *For a sequence* $A_1, A_2, \cdots, \epsilon \mathbf{M}$ *we have* $\lim_{n\to\infty}$ strong $A_n = 0$ *if and only if every subsequence* $A_{\bar 1}, A_{\bar 2}, \cdots$ *of* $A_1, A_2, \cdots$ *possesses a subsequence which is a* Σ*-sequence.*

PROOF: Necessity: $\lim_{n\to\infty}$ strong $A_n = 0$ implies for every subsequence $A_{\bar 1}, A_{\bar 2}, \cdots$ that $\lim_{n\to\infty}$ strong $A_{\bar n} = 0$. Hence Lemma 4 applies to $A_{\bar 1}, A_{\bar 2}, \cdots$ and so it has a subsequence $A_{1'}, A_{2'}, \cdots$ which is a Σ-sequence.

Sufficiency: Assume that we do not have $\lim_{n\to\infty}$ strong $A_n = 0$. Then there exists an $f \epsilon \mathfrak{H}$ for which we do not have $\lim_{n\to\infty} \| A_n f \| = 0$. Consequently there exists a subsequence $A_{\bar 1}, A_{\bar 2}, \cdots$ of $A_1, A_2, \cdots$ with $\lim_{n\to\infty} \| A_{\bar n} f \| = \alpha$ for an $\alpha \neq 0$ (but possibly $\alpha = \infty$). Thus for every subsequence, $A_{1'}, A_{2'}, \cdots$ of $A_{\bar 1}, A_{\bar 2}, \cdots$ equally $\lim_{n\to\infty} \| A_{n'} f \| = \alpha$ thus excluding $\lim_{n\to\infty}$ strong $A_{n'} = 0$. Hence Lemma 3 excludes that $A_{\bar 1}, A_{\bar 2}, \cdots$ be a Σ sequence.

Therefore, if the condition of our lemma is satisfied, we must have $\lim_{n\to\infty}$ strong $A_n = 0$.

Replacing $A_1, A_2, \cdots$ by $A_1 - A, A_2 - A, \cdots$ we can conclude from Lemma 5,

THEOREM II. *The notion* (γ) *(strong convergence) is purely algebraical.*

§6. (δ) can be deduced from (γ) by a known argument, which, nevertheless, we will give in full for the sake of completeness.

LEMMA 6. *If* $\lim_{n\to\infty}$ weak $A_n = A$ *there exists a subsequence* $A_{1'}, A_{2'}, \cdots$ *of* $A_1, A_2, \cdots$ *such that* $\lim_{n\to\infty}$ strong $\frac{1}{n} \sum_{m=1}^{n} A_{m'} = A$.

PROOF. Since we may replace $A, A_1, A_2, \cdots$ by $0, A_1 - A, A_2 - A, \cdots$ there is no loss of generality if we assume that $A = 0$, i.e. $\lim_{n\to\infty}$ weak $A_n = 0$.

By [1], p. 382, footnote 35, $||| A_1 |||, ||| A_2 |||, \cdots$ is bounded. Let α be a bound, i.e. $||| A_n ||| \leqq \alpha < \infty$.

Consider now an everywhere dense sequence $f_1^0, f_2^0, \cdots$ in $\mathfrak{H}$.

We shall define a subsequence $1', 2', \cdots$ of $1, 2, \cdots$ by induction. Assume therefore that $1', 2', \cdots, (m-1)'$ are already defined. We shall now define m'.

Since $\lim_{n\to\infty}$ weak $A_n = 0$, $\lim_{n\to\infty} (A_n f, g) = 0$ for any two f and $g \epsilon \mathfrak{H}$. Hence, in particular $\lim_{n\to\infty} (A_n f_i^0, A_l f_i^0) = 0$ for all $l = 1', 2', \cdots, (m-1)'$ and all $i = 1, \cdots, m$. Consequently there exists a $k(m)$ such that $n \geqq k(m)$ implies $| (A_n f_i^0, A_{l'} f_i^0) | \leqq 1/2^m$ for all $l' = 1', \cdots, (m-1)'$ and all $i = 1, \cdots, m$. Now choose $m' \geqq k(m)$ and $> (m-1)'$.

Thus $1' < 2' < \cdots$ and $|(A_{m'}f_i^0, A_{l'}f_i^0)| \leqq 1/2^m$ for $m > l$ and $m \geqq i$. Interchanging m, l [3] gives $|(A_{m'}f_i^0, A_{l'}f_i^0)| \leqq 1/2^l$, for $m < l$ and $l \geqq i$. Summing up:

$$\text{(2)} \qquad \{|(A_{m'}f_i^0, A_{l'}f_i^0)| \leqq 1/2^{\text{Max}(m,l)} \quad \text{if } m \neq l \text{ and Max } (m, l) \geqq i.$$

Now $n \geqq i$ yields

$$\begin{aligned}
\|(\textstyle\sum_{m=1}^n A_{m'})f_i^0\|^2 &= (\textstyle\sum_{m=1}^n A_{m'}f_i^0, \sum_{m=1}^n A_{m'}f_i^0) \\
&= \textstyle\sum_{m,l=1}^n (A_{m'}f_i^0, A_{l'}f_i^0) \\
&= \textstyle\sum_{m,l=1}^{i-1} (A_{m'}f_i^0, A_{l'}f_i^0) + \sum_{m=i}^n (A_{m'}f_i^0, A_{m'}f_i^0) \\
&\qquad + \textstyle\sum_{m,l=1(m\neq l,\text{Max}(m,l)\geqq i)}^n (A_{m'}f_i^0, A_{l'}f_i^0) \\
&\leqq (i-1)^2\alpha^2\|f_i^0\|^2 + (n-i+1)\alpha^2\|f_i^0\|^2 \\
&\qquad + \textstyle\sum_{m,l=1(m\neq l,\text{Max}(m,l)\geqq i)}^n 1/2^{\text{Max}(m,l)} \\
&= \alpha^2 \cdot \|f_i^0\|^2(n + (i-1)(i-2)) + \textstyle\sum_{k=i}^n \frac{2k-2}{2^k}{}^{4} \\
&\leqq \alpha^2\|f_i^0\|^2(n + (i-1)(i-2)) + 3^{5}
\end{aligned}$$

and so

$$\left\|\left(\frac{1}{n}\sum_{m=1}^n A_{m'}\right)f_i^0\right\| \leqq \frac{1}{n}\sqrt{\|f_i^0\|^2\alpha^2[n + (i-1)(i-2)] + 3}$$

This implies $\lim_{n\to\infty}\left\|\left(\frac{1}{n}\sum_{m1}^n A_{m'}\right)f_i^0\right\| = 0.$

Thus we have shown: The set $\mathfrak{S}$ of all $g \in \mathfrak{H}$ with

$$\lim_{n\to\infty}\left\|\left(\frac{1}{n}\sum_{m=1}^n A_{m'}\right)g\right\| = 0 \quad \text{contains all } f_1^0, f_2^0, \cdots.$$

Hence $\mathfrak{S}$ is everywhere dense in $\mathfrak{H}$. Since all $|||A_{m'}||| \leqq \alpha \left|\left|\left|\frac{1}{n}\sum_{m=1}^n A_{m'}\right|\right|\right| \leqq \alpha$ for $n = 1, 2, \cdots$. Thus the $\frac{1}{n}\sum_{m=1}^n A_{m'}$ are uniformly bounded and consequently $\mathfrak{S}$ is a closed set. So $\mathfrak{S} = \mathfrak{H}$ and $\lim_{n\to\infty}$ strong $\frac{1}{n}\sum_{m=1}^n A_{m'} = 0$.

This completes the proof.

LEMMA 7. *$\lim_{n\to\infty} A_n = A$ if and only if every subsequence $A_{\text{I}}, A_{\text{II}}, \cdots$ of $A_1, A_2, \cdots$ possesses a subsequence $A_{1'}, A_{2'}, \cdots$ such that*

$$\lim_{n\to\infty} \text{strong } \frac{1}{n}\sum_{m=1}^n A_{m'} = A.$$

[3] Remember that $(A_{m'}f_i^{(0)}, A_{l'}f_i^{(0)}) = (A_{l'}f_i^{(0)}, A_{m'}f_i^{(0)})$.

[4] We introduce a new summation variable, $k = \text{Max}(i, j)$. Given k the number of paris i, j which belong to it is clearly $2k - 2$.

[5] We have $\sum_{k=1}^\infty \frac{2k-2}{2^k} \leqq \sum_{k=1}^\infty \frac{2k-1}{2^k} = 3$.

PROOF. Necessity: Assume $\lim_{n\to\infty}$ weak $A_n = A$. For every subsequence $A_{\bar{1}}, A_{\bar{2}}, \cdots$ of $A_1, A_2, \cdots$, $\lim_{n\to\infty}$ weak $A_{\bar{n}} = A$. Hence Lemma 6 applies to $A_{\bar{1}}, A_{\bar{2}}, \cdots$ and so it has a subsequence $A_{1'}, A_{2'}, \cdots$ such that $\lim_{n\to\infty}$ strong $\frac{1}{n}\sum_{m=1}^{n} A_{m'} = A$. Thus the necessity of the hypothesis is established.

Sufficiency: Assume that the hypothesis is true but $\lim_{n\to\infty}$ weak $A_n = A$ is not true. Then there exists an f and $g \in \mathfrak{H}$ for which we do not have $\lim_{n\to\infty} (A_n f, g) = (Af, g)$. Consequently there exists a subsequence $A_{1^*}, A_{2^*}, \cdots$ of $A_1, A_2, \cdots$ with $\lim_{n\to\infty} (A_{n^*}f, g) = \alpha$ for an $\alpha \neq (Af, g)$ (but possibly $\alpha = \infty$). Hence if $A_{1'}, A_{2'}, \cdots$ is any subsequence of $A_{1^*}, A_{2^*}, \cdots$

$$\lim_{n\to\infty} (A_{n'}f, g) = \alpha \quad \text{and} \quad \alpha = \lim_{n\to\infty} \frac{1}{n}\sum_{m=1}^{n} (A_{m'}f, g)$$

$$= \lim_{n\to\infty} \left(\frac{1}{n}\sum_{m=1}^{n} A_{m'}f, g\right).$$

On the other hand, by hypothesis there is a subsequence $A_{1'}, A_{2'}, \cdots$ of $A_{1^*}, A_{2^*}, \cdots$ such that $\lim_{n\to\infty}$ strong $\frac{1}{n}\sum_{m=1}^{n} A_{m'} = A$ and consequently $\lim_{n\to\infty} \left(\frac{1}{n}\sum_{m=1}^{n} A_{m'}f, g\right) = (Af, g) \neq \alpha$. This contradicts the last result of the preceding paragraph, which states that for every such sequence, the limit is α.

This contradiction proves the sufficiency of the hypothesis.

With Lemma 7, we have shown

THEOREM III'. *The notion* (δ) (*weak convergence*) *is purely algebraical.*

INSTITUTE FOR ADVANCED STUDY

Annals of Mathematics
Vol. 44, No. 4, October, 1943

ON RINGS OF OPERATORS. IV

By F. J. Murray and J. von Neumann

Received February 24, 1943

INTRODUCTION

§1. The object of this paper is the investigation of the isomorphism properties of those operator rings which are *factors* and which are the substratum of several earlier publications of the authors. (Cf. [5], [6] and [8]. For the detailed definitions and also for the cases of factors—of which there will be more to say below—cf. [5], p. 138, Definition 3.1.2, p. 172, Theorem VIII.) The discrete cases—i.e. the cases (I)—have been exhaustively dealt with before. (Cf. [5], p. 173, Lemma 8.6.1, p. 139, Definitions 3.2.1 and 3.2.2 and Lemmas 3.2.1–3.2.3.) The purely infinite case—i.e. the case (III)—is the most refractory of all and we have, at least for the time being, scarcely any tools to investigate it. ([8] deals mainly with these factors.) Thus we are left with the continuous cases—i.e. the cases (II)—and they are our main objective in this paper.

An added justification of this program may be found in the fact that among all factors those of the finite continuous case—case (II_1)—have the strongest immediate interest. (Cf. [5], part V, [6] Chapter IV and Appendix.)

It will be seen however that the discrete cases can be included in our discussion with scarcely any extra effort. So we shall deal with them, i.e. with all cases but the purely infinite one.

For the discrete and continuous cases, the finite ones—i.e. the cases (I_n) $(n = 1, 2, \cdots)$ and (II_1) respectively—are basic because the infinite ones—i.e. the cases (I_∞) and (II_∞) respectively—can be subsequently described with their help. (Cf. Theorem IX and Lemma 3.1.6) Therefore we shall direct our main effort on the finite cases. And since the discrete ones—(the cases (I_n), $(n = 1, 2, \cdots)$)—are just the finite order matrix rings, this means essentially the above-mentioned continuous finite case (II_1).

§2. Let us now state the main problems of isomorphism more precisely.

Consider an operator ring $\mathbf{M}$ in a Hilbert space $\mathfrak{H}$ which contains 1. (We do not restrict $\mathbf{M}$ in any other way yet. For the significant definitions, cf. [2], pp. 388–389, Definitions 1–3). Then there exist two kinds of notions in $\mathbf{M}$: First, those which can be expressed in terms of the entity 1 (the unit operator) and the operations αA (α any complex number), A^*, $A + B$, $A \cdot B$ alone and referring only to the operators belonging to $\mathbf{M}$; Second those which need other things as well, e.g. operators outside of $\mathbf{M}$, elements of $\mathfrak{H}$, etc. The former notions are *purely algebraical*, the latter ones are not; we shall also call them spatial.

These notions were already investigated by the second author in [9].

It seems worth while to formulate this distinction in terms of isomorphisms. Let, for each $i = 1, 2$ a Hilbert space $\mathfrak{H}_i$ and an operator ring $\mathbf{M}_i$ in $\mathfrak{H}_i$ be given.

We now define

A) Spatial isomorphism of $\mathbf{M}_1$ and $\mathbf{M}_2$. This is a one-to-one isomorphic mapping of $\mathfrak{H}_1$ on $\mathfrak{H}_2$ (i.e. one which is linear and isometric—unitary) which carries $\mathbf{M}_1$ into $\mathbf{M}_2$.

B) Algebraical ring isomorphism of $\mathbf{M}_1$ and $\mathbf{M}_2$. This is a one-to-one mapping of $\mathbf{M}_1$ onto $\mathbf{M}_2$ which leaves the entity 1 and the operations αA (α any complex number) A^*, $A + B$, AB (when passing from $\mathbf{M}_1$ to $\mathbf{M}_2$) invariant.

(Cf. also [5], p. 145, §5.2, in particular Definition 5.2.1. We have added the requirement that 1 be invariant.)

Any spatial property is invariant under the *spatial isomorphisms* of A), while the purely algebraical properties are characterized by their invariance under the *algebraical isomorphisms* of B).

Clearly A) implies B) while the converse need not be true.

An important ring theoretical notion which is only spatial is $\mathbf{M}'$. (Cf. [2], p. 388, Def. 3. For the spatial character, cf. the detailed discussion of §3.3.) Nevertheless the factor property, i.e. $\mathbf{M}\cdot\mathbf{M}' = (\alpha\cdot 1)$ (cf. §3.3, loc. cit., and also the beginning of §1) is purely algebraical since it states that the operators $\alpha\cdot 1$ exhaust the *center* of $\mathbf{M}$ (Cf. [5], p. 138, Def. 3.1.2.)

Now our program is subdivided as follows.

Question I: When does B) hold?

Question II: Under what additional conditions does B) imply A)?

Question II was already investigated in a special case in [6]. It was shown there that under certain conditions A) and B) are equivalent. (Cf. [6], p. 244, Theorem XI.) We shall obtain a complete answer to Question II in Theorem X.

Question I is more difficult. It coincides with Problem 6 in [5], p. 172, and the present paper contains what progress the authors have been able to make in that direction. The main results are these: An extensive class of factors of case (II_1) which are all isomorphic to each other in the sense B) will be determined. These factors are called "approximately finite". (Cf. Theorems XII, XIV, which are based on Defs. 4.1.1, 4.3.1, 4.5.2 and 4.6.1 below.) The isomorphisms announced in [5], p. 229, (v) are contained in this class. On the other hand, certain factors of case (II_1) which are not isomorphic to the approximately finite ones will be constructed. (Cf. Theorems XVI, XVI'.)

§3. There are indications to the effect that the approximately finite factors are the simplest among those of case (II_1) but the evidence is not quite conclusive. It is true that every factor in case (II_1) has an approximately finite subring. (Cf. Theorem XIII). However, this "imbedding theorem" does not settle the matter, since the analogue of the Cantor-Bernstein "equivalence theorem" is not true: Two factors in the case (II_1) might be such that each is isomorphic to a sub-ring of the other, but the factors themselves may not be isomorphic. (An example of this is given in the appendix.) Hence the possibility exists that any factor in the case (II_1) is isomorphic to a sub-ring of any other such factor.

§4. The best we can do at present concerning the isomorphism problem in the case (II_1) is this: The limits of the approximate finite case are extended rather far in §5.2 and §5.6. The existence of non-approximately-finite factors in this case is established in Theorems XVI, XVI′. Certain algebraical invariants of factors in the case (II_1) are formed. ((1), (2), in §4.6, and the property Γ in Def. 6.1.1) of which the first two are probably of greater general significance, but the last one has so far been put to greater practical use. (Cf. the remark at the beginning of §6.1.)

The isomorphism questions of the case (II_∞) are reduced to those of the case (II_1). (This follows from Theorem IX.) Those of the discrete cases—the cases (I)—have been settled before (cf. above at the beginning of §1; also α) in Theorem IX).

This enumeration exhausts our present program. It answers Problems 6, 7 in [5], p. 172, and [v] eod., p. 229, as far as possible at this moment.

§5. We add that §5.3 contains a new technique of constructing factors in the case (II_1). This seems to be considerably simpler than our previous procedures, but it is closely related to them. (Cf. §5.4, §5.5.) It also throws some light on the meaning of this work from the point of view of the unitary representation theory of groups (cf. the remarks after Lemma 5.3.4). Further generalizations are probably possible and important. (Cf. the remark at the end of §5.6.)

The result that not all factors in the case (II_1) are isomorphic to each other, expressed in Theorem XVI′ deserves some further comment. From the point of view of the systematic build-up of the paper, it seemed best to put it at the end. It is, however, intelligible and of interest in itself, and it can be derived independently of most of the paper. The reader who is primarily interested in this particular result need only read §5.3, Def. 6.11, Lemma 6.1.1, Lemma 6.2.1, Lemma 6.2.2. The entire remainder of this paper is unnecessary for this purpose, in particular the extensive theories of algebraical and spatial types, of genera, and of approximate finiteness and its various equivalent forms. But we think, nevertheless, that knowledge of the whole paper will help to see this result too in the proper perspective.

§6. The notations are the same as in the papers referred to in the bibliography, particularly [5] and [6].

We use the Kronecker-Weierstrass symbol in the general sense:

$$\delta_{a,b} = \begin{cases} 1 \text{ for } a = b \\ 0 \text{ otherwise.} \end{cases}$$

for any objects a, b.

Bibliography

[1] John von Neumann, *Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren*, Math. Annalen, vol. 102 (1929), pp. 49–131.

[2] John von Neumann, *Zur Algebra der Funktionaloperatoren*, Math. Annalen, vol. 102 (1929), pp. 370–427.

[3] JOHN VON NEUMANN, *Über Funktionen von Funktionaloperatoren*, Annals of Math., vol. 32 (1931), pp. 191–226.
[4] JOHN VON NEUMANN, *On a certain topology for rings of operators*, Annals of Math., vol. 37 (1936), pp. 111–115.
[5] F. J. MURRAY AND JOHN VON NEUMANN, *On rings of operators*, Annals of Math., vol. 37 (1936), pp. 116–229.
[6] F. J. MURRAY AND JOHN VON NEUMANN, *On rings of operators II*, Trans. Amer. Math. Socl, vol. 41 (1937), pp. 208–248.
[7] JOHN VON NEUMANN, *On infinite direct products*, Compositio Mathematica, vol. 6 (1938), pp. 1–77.
[8] JOHN VON NEUMANN, *On rings of operators III*, Annals of Math, vol. 41 (1940), pp. 94–161.
[9] JOHN VON NEUMANN, *On some algebraical properties of operator rings*, Annals of Math., vol. 44, preceding immediately.

CONTENTS

CHAPTER I. THE METRIC OF M

PURELY ALGEBRAICAL CHARACTER OF VARIOUS NOTIONS

§1.1 Consider a Hilbert space $\mathfrak{H}$ and an operator ring $\mathbf{M}$ in $\mathfrak{H}$. We know from §2 in the introduction as well as from [9] that the purely algebraical character of certain notions in $\mathbf{M}$ is neither obvious nor certain.

It was demonstrated in [9], however, that the following notions in $\mathbf{M}$ are purely algebraical.

Definiteness of an $A \in \mathbf{M}$; the numerical value $|||A|||$ of the bound of an $A \in \mathbf{M}$; convergence in the sense of the strong and of the weak operator topologies. (Convergence in the sense of the strongest topology is the same as in the strong one. Cf. [4], p. 112, end of §2.)

We did not establish the same thing for the operator topologies themselves,

i.e. for the strongest, the strong and the weak. (For the definitions, cf. [4], p. 111, §2, and [2], pp. 381–382.) The uniform topology, (cf. [2], pp. 384–386) causes no difficulty, since $|||\, A\, |||$ is purely algebraical, but it will play no role in our discussions. This is very unsatisfactory since the notion of a subring of **M** must be based on one of these topologies. (Cf. [4], p. 112, beginning of §3; [2], p. 396, end of §3; p. 388, Definition 1. Best consulted in the reverse order.) They cannot be replaced by convergence notions, nor by the uniform topology. (Cf. in particular [2], p. 382–3 and p. 384.) And yet the notion of a subring ought to be purely algebraical.

For our present purposes it will be adequate to settle these questions for a restricted class of rings **M** and this we now do.

§1.2 Observe first that the following notions are purely algebraical.

The factor character of **M** and the case to which **M** belongs, the numerical value of the relative dimension function $D_{\mathbf{M}}(E)$ (E any projection $\epsilon\, \mathbf{M}$) apart from its normalization and even the standard normalization in the cases where one is defined; the numerical value of the relative trace $Tr_{\mathbf{M}}(A)$, ($A\, \epsilon\, \mathbf{M}$) in the finite cases.

These assertions were established in [5], p. 145, Lemma 5.2.1; p. 173, Theorem IX; [6], p. 219, Property IV (since that characterization is purely algebraical).

We assume now that **M** is a factor in a finite case.[1] We shall use the relative dimension function $D_{\mathbf{M}}(E)$ and it will be advantageous not to normalize it. We shall also use the relative trace $Tr_{\mathbf{M}}(A)$ with its usual properties. We define for every $A\, \epsilon\, \mathbf{M}$

$$[[A]] = \sqrt{Tr_{\mathbf{M}}(A^*A)} = \sqrt{Tr_{\mathbf{M}}(AA^*)}. \tag{1.2.α}$$

(Cf. [6], p. 241, Lemma 4.3.2. The considerations of [8], p. 102, Def. 1.3.1 and Theorem II, are somewhat more general but of the same type.) Owing to our above remarks, the numerical value of $[[A]]$ is purely algebraical too.

By [6], p. 235, Theorem III, there exists a fixed finite system $g_1, \cdots, g_m\, \epsilon\, \mathfrak{H}$ (depending on **M** only), such that for all $A\, \epsilon\, \mathbf{M}$

$$Tr_{\mathbf{M}}(A) = \textstyle\sum_{i=1}^{m}(Ag_i\,, g_i). \tag{1.2.β}$$

Combining (1.2.α) with (1.2.β), and remembering that $(A^*Ag,\, g) = \|\, Ag\, \|^2$, $(AA^*g,\, g) = \|\, A^*g\, \|^2$ we obtain

$$[[A]] = \sqrt{\textstyle\sum_{i=1}^{m} \|\, Ag_i\, \|^2} = \sqrt{\textstyle\sum_{i=1}^{m} \|\, A^*g_i\, \|^2}. \tag{1.2.γ}$$

This formula is useful for many purposes, although it obscures the purely algebraical character of $[[A]]$.

$[[A]]$ has all the properties of a *norm* in a linear space, and accordingly $[[A - B]]$ has all those of a *distance*. (Cf. for these and for some further properties [[6]], p. 242, Property II°. We need these for **M** only, and not, as loc. cit., for

[1] A finite case is either a (I_n), $n = 1, 2, \cdots$ or a (II_1). The inclusion of the former is unnecessary, since our real interest is in case (II_1). But these considerations have the totality of all finite cases as their natural scope.

its extension $Q(\mathbf{M})$. All these properties can be read off the representation (1.2.γ)—the decisive fact being, of course, that we have them both.)

Thus $\mathbf{M}$ has become a topological space in a new way, with a purely algebraical metric.[2]

We now define

DEFINITION 1.2.1. *A sequence* $A_1, A_2, \cdots \in \mathbf{M}$ *is metrically convergent to* $A \in \mathbf{M}$ *if* $\lim_{\nu\to\infty} [[A_\nu - A]] = 0$. *It is restrictedly metrically convergent to* $A \in \mathbf{M}$ *if* $\lim_{\nu\to\infty} [[A_\nu - A]] = 0$ *and the (numerical) sequence* $|||\, A_1 \,|||, |||\, A_2 \,|||, \cdots$ *is bounded.*

The notions of closure for subsets $\mathbf{S}$ of $\mathbf{M}$ which are induced by these two notions of convergence, are the "*metrical closure*" and the "*restricted metrical closure*" of $\mathbf{S}$.[3]

We shall also need:

DEFINITION 1.2.2. *A subset* $\mathbf{S}$ *of* $\mathbf{M}$ *is linear if* $A, B \in \mathbf{S}$ *imply* $\alpha A, A + B \in \mathbf{S}$. *It is an algebra if* $A, B \in \mathbf{S}$ *imply* $\alpha A, A^*, A + B, AB \in \mathbf{S}$.

Clearly both these definitions define purely algebraical notions.

Now a subring $\mathbf{S}$ of $\mathbf{M}$ is a closed algebra, where closure may be understood in any one of the three operator topologies: the strongest, the strong, or the weak (cf. the end of §1.1). We shall establish the purely algebraical character of the notion of a subring by proving the equivalence of the notion of closure in all these topologies to the purely algebraical closures of Def. 1.2.1 for a set $\mathbf{S}$ which is an algebra.[4]

§1.3 Observe first:

LEMMA 1.3.1. *For the* $g_1, \cdots, g_m$ *of* (1.2.β) *and* (1.2.γ) *the* $A'g_i$, $(A' \in \mathbf{M}'$, $i = 1, \cdots, m)$ *span the closed linear set* $\mathfrak{H}$.

PROOF. For a fixed i $(= 1, \cdots, m)$ the $A'g_i$, $A' \in \mathbf{M}'$ span the closed linear set $\mathfrak{M}_{g_i}^{\mathbf{M}'}$ (Cf. [5], p. 143, Def. 5.1.1.) Put $\mathfrak{N} = [\mathfrak{M}_{g_i}^{\mathbf{M}'}, i = 1, \cdots, m]$. We must prove $\mathfrak{N} = \mathfrak{H}$. Now $\mathfrak{M}_{g_i}^{\mathbf{M}'} \;\eta\; \mathbf{M}$ (cf. loc. cit., Lemma 5.1.1). Hence $\mathfrak{N} \;\eta\; \mathbf{M}$ and $E = P_{\mathfrak{N}} \in \mathbf{M}$. Also $g_i \in \mathfrak{M}_{g_i}^{\mathbf{M}'} \subset \mathfrak{N}$ so

$$Eg_i = g_i. \tag{1.3.α}$$

Now (1.2.γ) shows that if $A \in \mathbf{M}$ and $Ag_i = 0$ for $i = 1, \cdots, m$, then $A = 0$. Apply this to $A = 1 - E$. Then (1.3.α) yields $1 - E = 0$, $P_{\mathfrak{N}} = E = 1$ as desired.

---

[2] The uniform topology, already mentioned in §1.1, is actually of the same character: $|||\, A \,|||$, $|||\, A - B \,|||$ are a norm and a metric, just as $[[A]]$, $[[A - B]]$. But only the latter is useful in the theory of operator rings. This is due to its connection with the strong topology, which will appear in §1.4 and §1.5.

[3] The metrical closure is the conventional topological notion, based on the specified metric. The restricted metrical closure is not of that nature. Thus it is not evident that the set of all limits of all restrictedly metrically convergent sequences from a given set $\mathbf{S}$ is necessarily closed in that sense. Actually this is not true for all sets $\mathbf{S}$ although when $\mathbf{S}$ is an algebra it does follow from the results of §1.5 (cf. there).

[4] Observe that restrictions on both the ring $\mathbf{M}$ and its subset $\mathbf{S}$ are necessary in order to secure this result. $\mathbf{M}$ must be a factor in a finite case, $\mathbf{S}$, an algebra.

We now are able to prove:

LEMMA 1.3.2. *The sequence $A_1, A_2, \cdots \in \mathbf{M}$ is strongly convergent to $A \in \mathbf{M}$ if and only if it is restrictedly metrically convergent to this A.*

PROOF. Necessity: Assume that $A_1, A_2, \cdots$ is strongly convergent to A. Then the (numerical) sequence $|||A_1|||, |||A_2|||, \cdots$ is bounded by [2], p. 382, footnote 35, and $\lim_{\nu\to\infty} [[A_\nu - A]] = 0$ by (1.2.γ). Hence we have restricted metric convergence.

Sufficiency: Assume that $A_1, A_2, \cdots$ are restrictedly metrically convergent to A. We must show that they converge strongly too, i.e. that the set $\mathfrak{M}$ of all f with $\lim_{\nu\to\infty} \|(A_\nu - A)f\| = 0$ is $\mathfrak{H}$. $\mathfrak{M}$ is obviously a linear set. Since the (numerical) sequence $|||A_1|||, |||A_2|||, \cdots$ is bounded, $\mathfrak{M}$ is also closed. Hence it suffices to show that $\mathfrak{M}$ spans the closed linear set $\mathfrak{H}$.

Now by (1.2.γ), $\|(A_\nu - A)g_i\| \leqq [[A_\nu - A]]$. Hence $\lim_{\nu\to\infty} \|(A_\nu - A)g_i\| = 0$. Therefore for every $A' \in \mathbf{M}'$, $\lim_{\nu\to\infty} \|A'(A_\nu - A)g_i\| = 0$. But $A'(A_\nu - A) = (A_\nu - A)A'$ since $A' \in \mathbf{M}'$ and $A_\nu, A \in \mathbf{M}$. Consequently $\lim_{\nu\to\infty} \|(A_\nu - A)A'g_i\| = 0$ and all $A'g_i \in \mathfrak{M}$. Hence $\mathfrak{M}$ spans the closed linear set $\mathfrak{H}$ by Lemma 1.3.1, and the proof is completed.

Thus closure with respect to strong convergence is equivalent to restricted metric closure for every subset $\mathbf{S}$ of $\mathbf{M}$. This however is not enough to establish the purely algebraical character of the notion of a subring, since that involves topological closure which cannot be replaced by convergence closure (cf. the end of §1.1). In fact, from that point of view we have made no progress at all, since we knew all along from [9] that strong convergence is a purely algebraical notion for all rings $\mathbf{M}$. Besides our present restrictions on $\mathbf{M}$, we shall also have to restrict $\mathbf{S}$ to algebras before we can get any further (cf. the end of §1.2).

§1.4 Consider first the following two simple results, valid for all subsets $\mathbf{S}$ of $\mathbf{M}$.

LEMMA 1.4.1. *Strong closure of* $\mathbf{S}$ *implies the restricted metric closure.*

PROOF. Strong closure of $\mathbf{S}$ implies closure with respect to strong convergence, and by Lemma 1.3.2 the latter is equivalent to restricted metric closure.

LEMMA 1.4.2. *Metric closure of* $\mathbf{S}$ *implies its strong closure.*

PROOF. Let $\mathbf{S}$ be metrically closed and consider a strong condensation point A of $\mathbf{S}$. Since $\mathbf{S} \subset \mathbf{M}$ and $\mathbf{M}$ is strongly closed, this implies $A \in \mathbf{M}$.

For every $\nu = 1, 2, \cdots$ choose an $A_\nu \in \mathbf{S}$ with $\|(A_\nu - A)g_i\| \leqq 1/\nu$ for $i = 1, \cdots, m$. This can be done since, by the definition of a limit point, we must have at least one $A_\nu \in \mathbf{M}$ in the strong neighborhood $U_3(A, g_1, \cdots g_m; 1/\nu)$ of A (cf. [2], §2, p. 381). Then (1.2.γ) gives $[[A_\nu - A]] \leqq \sqrt{m}/\nu$ and hence $\lim_{\nu\to\infty} [[A_\nu - A]] = 0$. Thus $A_1, A_2, \cdots$ converves metrically to A and since $\mathbf{S}$ is metrically closed, $A \in \mathbf{S}$. This completes the proof.

Now if we can show that the restricted metric closure implies metric closure, the above lemmas will yield the equivalence of the strong, the metric and the restricted metric notions of closure. As pointed out at the end of §1.3, this is what we need. The assumption that $\mathbf{S}$ is an algebra will be needed to secure the above implication.[4]

§1.5 In what follows we shall make repeated use of the notion and properties of the functions of bounded Hermitian and of unitary operators,[5] and it will be convenient to omit specific references to that theory.

LEMMA 1.5.1. (i) *If A is bounded Hermitian, then $(A - i1)^{-1}$ can be formed and*

$$\frac{A + i1}{A - i1} = (A - i\cdot 1)^{-1}(A + i1) = (A + i1)(A - i1)^{-1}$$

is unitary.

(ii) *If A, $B \in \mathbf{M}$ and Hermitian, then $\frac{A + i1}{A - i1}, \frac{B + i1}{B - i1} \in \mathbf{M}$ too, and*

$$\left[\left[\frac{A + i1}{A - i1} - \frac{B + i1}{B - i1}\right]\right] \leqq 2[[A - B]].$$

PROOF. Ad. (i). Consider the two functions

$$f(\lambda) = \frac{1}{\lambda - i}, \qquad g(\lambda) = \frac{\lambda + i}{\lambda - i}.$$

Since they are continuous for all real λ we can form $f(A)$ and $g(A)$. We also have

$$(\lambda - i)f(\lambda) = f(\lambda)(\lambda - i) = 1$$
$$(\lambda + i)f(\lambda) = f(\lambda)(\lambda + i) = g(\lambda).$$

Therefore $f(A) = (A - i1)^{-1}$ and $g(A) = (A + i1)f(A) = f(A)(A + i1)$. Thus $g(A) = (A + i1)(A - i1)^{-1} = (A - i1)^{-1}(A + i1)$ or $g(A) = (A + i1)/(A - i1)$ in the sense indicated above.

Since $\overline{g(\lambda)}g(\lambda) = 1 = g(\lambda)\overline{g(\lambda)}$ for all real λ, $g(A)^*g(A) = 1 = g(A)g(A)^*$ and hence $g(A)$ is unitary. Thus the proof of (i) is completed. We note also that inasmuch as $|f(\lambda)| \leqq 1$ for all real λ, $|||f(A)||| \leqq 1$ or

$$(1.5.\alpha) \qquad |||(A - i1)^{-1}||| \leqq 1.$$

Ad (ii), A, $B \in \mathbf{M}$ imply $g(A)$, $g(B) \in \mathbf{M}$ or $(A + i1)/(A - i1)$, $(B + i1)/(B - i1) \in \mathbf{M}$.

We have

$$\begin{aligned}\frac{A + i1}{A - i1} - \frac{B + i1}{B - i1} &= (A - i1)^{-1}(A + i1) - (B + i1)(B - i1)^{-1} \\ &= (A - i\cdot 1)^{-1}\{(A + i1)(B - i1) \\ &\qquad - (A - i\cdot 1)(B + i\cdot 1)\}(B - i1)^{-1} \\ &= -2i(A - i1)^{-1}(A - B)(B - i\cdot 1)^{-1}.\end{aligned}$$

[5] Cf., e.g., [3], pp. 205 and 220. The real function-theory methods of [3] are actually much more than what is needed for the limited objectives of §1.5, as only continuous functions (sometimes even polynomials) of operators occur. ((i) in Lemma 1.5.1 can even be proven by elementary evaluations.) But it would take up too much space to go into such methodological questions systematically.

Consequently

$$\left[\left[\frac{A+i1}{A-i1}-\frac{B+i1}{B-i1}\right]\right] \leqq 2\,|||\,(A-i\cdot 1)^{-1}\,|||\,[[A-B]]\,|||\,(B-i\cdot 1)^{-1}\,|||$$

and by (1.5.α)

$$\left[\left[\frac{A+i1}{A-i1}-\frac{B+i1}{B-i1}\right]\right] \leqq 2[[A-B]].$$

LEMMA 1.5.2. *Let a function $\varphi(\xi)$ be given, defined and continuous for all ξ with $|\xi| = 1$. Then there exists a function $\omega_1(\epsilon)$ (given φ, ω_1 is determined) defined and >0 for all $\epsilon > 0$, with the following property: If U, V are unitary and ϵ **M** then $\varphi(U)$, $\varphi(V)$ ϵ **M** and $[[U - V]] < \omega_1(\epsilon)$ implies $[[\varphi(U) - \varphi(V)]] < \epsilon$.*

PROOF. That U, V unitary and ϵ **M** imply $\varphi(U)$ ϵ **M**, $\varphi(V)$ ϵ **M** is obvious.

We shall now show that it suffices to prove our assertion for all trigonometrical polynomials

$$\varphi(\xi) = \sum_{\nu=-n}^{n} a_\nu \xi^\nu.$$[6]

Assume accordingly that it is true for these and consider a general $\varphi(\xi)$.

Consider an $\epsilon' > 0$. Choose a trigonometrical polynomial

$$\varphi^{\epsilon'}(\xi) = \sum_{\nu=-n^{\epsilon'}}^{n^{\epsilon'}} a_\nu^{\epsilon'} \xi^\nu$$

such that $|\varphi(\xi) - \varphi^{\epsilon'}(\xi)| \leqq \frac{1}{3}\epsilon'$ for all ξ with $|\xi| = 1$[6]. Then the unitary of U, V implies $|||\,\varphi(U) - \varphi^{\epsilon'}(U)\,||| \leqq \frac{1}{3}\epsilon'$ and $|||\,\varphi(V) - \varphi^{\epsilon'}(V)\,||| \leqq \frac{1}{3}\epsilon'$. Hence, *a fortiori*,

$$[[\varphi(U) - \varphi^{\epsilon'}(U)]] \leqq \tfrac{1}{3}\epsilon', \qquad [\varphi(V) - \varphi^{\epsilon'}(V)]] \leqq \tfrac{1}{3}\epsilon'.$$

Denote the $\omega_1(\epsilon)$ of $\varphi^{\epsilon'}(\xi)$ by $\omega_1^{\epsilon'}(\epsilon)$. Then $[[U - V]] < \omega_1^{\epsilon'}(\frac{1}{3}\epsilon')$ implies

$$[[\varphi^{\epsilon'}(U) - \varphi^{\epsilon'}(V)]] < \tfrac{1}{3}\epsilon'$$

and so, owing to the above,

$$[[\varphi(U) - \varphi(V)]] < \epsilon'.$$

Thus $\omega_1(\epsilon') = \omega_1^{\epsilon'}(\frac{1}{3}\epsilon')$ meets our requirements for the original $\varphi(\xi)$.

It only remains therefore to prove our assertion for the trigonometrical polynomials

$$\varphi(\xi) = \sum_{\nu=-n}^{n} a_\nu \xi^\nu.$$

Obviously we may even restrict ourselves to the monomials

$$\varphi(\xi) = \xi^\nu, \qquad (\nu = 0 \pm 1, \pm 2, \cdots).$$

[6] Since $|\xi| = 1$ we may put $\xi = e^{i\alpha}$, α real, mod 2π. Now

$$\varphi(\xi) = \sum_{\nu=-n}^{n} a_\nu \xi^\nu = \sum_{\nu=-n}^{n} a_\nu e^{i\nu\alpha}.$$

Thus these $\varphi(\xi)$ are the trigonometrical polynomials with their familiar approximation properties. Cf. Zygmund, Antoni, "Trigonometrical Series", Warsaw (1935), §3.23 (iii), p. 47.

For $\nu = 0$ this is trivial, and for $\nu < 0$ the unitarity of U, V implies $[[U^\nu - V^\nu]] = [[(U^\nu - V^\nu)^*]] = [[U^{*\nu} - V^{*\nu}]] = [[U^{-\nu} - V^{-\nu}]]$.

Therefore we may even assume $\nu > 0$. In this case, however,

$$\begin{aligned} U^\nu - V^\nu &= \sum_{\mu=1}^{\nu}(U^\mu V^{\nu-\mu} - U^{\mu-1}V^{\nu-\mu+1}) \\ &= \sum_{\mu=1}^{\nu} U^{\mu-1}(U - V)V^{\nu-\mu}. \\ [[U^\nu - V^\nu]] &= [[\sum_{\mu=1}^{\nu} U^{\mu-1}(U - V)V^{\nu-\mu}]] \\ &= \sum_{\mu=1}^{\nu} ||| U^{\mu-1} ||| \, [[U - V]] \, ||| V^{\nu-\mu} ||| \\ &= \sum_{\mu=1}^{\nu} [[U - V]] = \nu[[U - V]]. \end{aligned}$$

Thus $\omega_1(\epsilon) = (1/\nu)\epsilon$ meets our requirements for these $\varphi(\xi)$.

LEMMA 1.5.3. *Let a function $\psi(\lambda)$ be given, defined and continuous for all real λ and for which $\lim_{\lambda\to\pm\infty}\psi(\lambda)$ exists.*[7] *Then there exists a function $\omega_2(\epsilon)$ (given ψ, ω is determined) defined and > 0 for all $\epsilon > 0$ with the following prôperty: If A, B are Hermitian and ϵ **M** then $\psi(A)$, $\psi(B) \,\epsilon$ **M** and $[[A - B]] < \omega_2(\epsilon)$ implies $[[\psi(A) - \psi(B)]] < \epsilon$.*

PROOF. That A, B Hermitian and ϵ **M** imply $\psi(A)$, $\psi(B) \,\epsilon$ **M** is obvious.

The correspondence

$$\xi = \frac{\lambda + i}{\lambda - i} \qquad \text{or equivalently} \qquad \lambda = i\,\frac{\xi + 1}{\xi - 1}$$

is a one-to-one and bicontinuous mapping of ξ with $|\xi| = 1$ on all real λ and $\lambda = \pm\infty$ where $\xi = 1$ corresponds to $\lambda = \pm\infty$. Hence

$$\varphi(\xi) = \begin{cases} \psi\left(i\,\dfrac{\xi + 1}{\xi - 1}\right) & \text{for} \quad |\xi| = 1, \xi \neq 1 \\ \cdot\lim_{\lambda\to\pm\infty}\psi(\lambda) & \text{for} \quad \xi = 1 \end{cases}$$

is defined and continuous for all ξ with $|\xi| = 1$. Clearly

$$\psi(\lambda) = \varphi\left(\frac{\lambda + i}{\lambda - i}\right).$$

We recall the $g(\lambda) = (\lambda + i)/(\lambda - i)$ used in the proof of Lemma 1.5.1 and let $U = g(A) = (A + i\cdot 1)/(A - i\cdot 1)$, $V = g(B) = (B + i1)/(B - i1)$. Then the above relations imply

$$\psi(A) = \varphi(U), \qquad \psi(B) = \varphi(V).$$

By Lemma 1.5.1, U and V are unitary and ϵ **M** and

$$[[U - V]] \leqq 2[[A - B]].$$

[7] We require $\lim_{\lambda\to\infty}\psi(\lambda) = \lim_{\lambda\to-\infty}\psi(\lambda)$. This condition could be removed, but it simplifies the proof and suffices for our present purposes.

Now Lemma 1.5.2 applies to $\varphi(\xi)$. Consider the resulting $\omega_1(\epsilon)$. Then $[[A - B]] < \frac{1}{2}\omega_1(\epsilon)$ implies $[[U - V]] < \omega_1(\epsilon)$ and hence $[[\varphi(U) - \varphi(V)]] < \epsilon$ or $[[\psi(A) - \psi(B)]] < \epsilon$. Therefore $\omega_2(\epsilon) = \frac{1}{2}\omega_1(\epsilon)$ meets our requirements.

LEMMA 1.5.4. *Let an algebra* $\mathbf{S}$ *in* $\mathbf{M}$ *be given and a sequence* $A_1, A_2, \cdots \in \mathbf{S}$ *which converges metrically to an* $A \in \mathbf{M}$. *Then there exists another sequence,* $A'_1, A'_2, \cdots \in \mathbf{S}$ *which converges restrictedly metrically to* A.

PROOF. $A_1, A_2, \cdots$ converges metrically to A; hence $A_1^*, A_2^*, \cdots$ converges metrically to A^*.[8] Consequently $(1/2)(A_1 + A_1^*), (1/2)(A_2 + A_2^*), \cdots$ converges metrically to $(1/2)(A + A^*)$, and $(1/2i)(A_1 - A_1^*), (1/2i)(A_2 - A_2^*), \cdots$ converges metrically to $(1/2i)(A - A^*)$. The operators occurring in the two last assertions are all Hermitian and $\in \mathbf{S}$. Also $A = (1/2)(A + A^*) + i(1/2i)(A - A^*)$. Now suppose our result has been established in the special case in which all the operators $A_1, A_2, \cdots$ are Hermitian. Then to treat the general case, we should apply the special result to the above-mentioned Hermitian "real and imaginary parts" and by taking the linear combination we should obtain the desired result since the notion of "restricted metrical convergence" holds for linear combinations of series if it holds for the series themselves. Thus we may assume in what follows that $A, A_1, A_2, \cdots$ are Hermitian operators.

Now form the function

$$\psi(\lambda)\begin{cases} = \lambda & \text{for } \quad |\lambda| \leqq |||A||| \\ = \dfrac{|||A|||^2}{\lambda} & \text{for } \quad |\lambda| \geqq |||A|||. \end{cases}$$

Lemma 1.5.3 applies to $\psi(\lambda)$ yielding $\omega_2(\epsilon)$.

Consider a $\nu = 1, 2, \cdots$. Choose a $\nu' = 1, 2, \cdots$ (depending on ν as well as on $A, A_1, A_2, \cdots$) with

$$[[A_{\nu'} - A]] < \omega_2\left(\frac{1}{\nu}\right).$$

Then Lemma 1.5.3 gives $[[\psi(A_{\nu'}) - \psi(A)]] < 1/\nu$. Since $\psi(\lambda) = \lambda$ for all λ with $|\lambda| \leqq |||A|||$, $\psi(A) = A$.[9] Thus we have

$$[[\psi(A_{\nu'}) - A]] < 1/\nu. \tag{1.5.β}$$

Form next the function

$$\theta(\lambda)\begin{cases} = 1 & \text{for } \quad |\lambda| \leqq |||A||| \\ = |||A|||^2/\lambda^2 & \text{for } \quad |\lambda| \geqq |||A|||. \end{cases}$$

It is clear that $\psi(\lambda) = \lambda \cdot \theta(\lambda)$. Choose a polynomial $p_\nu(\lambda)$ with

$$\theta(\lambda) \geqq p_\nu(\lambda) \geqq \text{Max}\,(\theta(\lambda) - 1/\nu \cdot |\lambda|, 0) \quad \text{for } \quad |\lambda| \leqq |||A_{\nu'}|||.$$[10]

[8] Since $[[A]] = [[A^*]]$, $A \to A^*$ is an isometric mapping.

[9] The set, $|\lambda| \leqq |||A|||$, contains the entire spectrum of A.

[10] This is an easy variant of the Weierstrass (polynomial) approximation theorem.

Then the polynomial $q_\nu(\lambda) = \lambda p_\nu(\lambda)$ has no constant term and we have

$$| q_\nu(\lambda) - \psi(\lambda) | \leqq 1/\nu \quad \text{for} \quad | \lambda | \leqq ||| A_{\nu'} |||.$$

Furthermore $0 \leqq q_\nu(\lambda) \leqq \psi(\lambda) \leqq ||| A |||$ and thus

$$| q_\nu(\lambda) | \leqq ||| A ||| \quad \text{for} \quad | \lambda | \leqq ||| A_{\nu'} |||.$$

These two inequalities imply

$$||| q_\nu(A_{\nu'}) - \psi(A_{\nu'}) ||| \leqq 1/\nu^{11}$$

and hence, *a fortiori*

$$(1.5.\gamma) \qquad [[q_\nu(A_{\nu'}) - \psi(A_{\nu'})]] \leqq 1/\nu$$

and

$$(1.5.\delta) \qquad ||| g_\nu(A_{\nu'}) ||| \leqq ||| A |||.^{11}$$

Combining (1.5.β) and (1.5.γ) gives

$$(1.5.\epsilon) \qquad [[q_\nu(A_{\nu'}) - A]] < 2/\nu.$$

Now put $A'_\nu = q(A_{\nu'})$. Then $A_{\nu'} \in \mathbf{S}$ implies $q_\nu(A_{\nu'}) \in \mathbf{S}$ since $q_\nu(\lambda)$ has no constant term. Hence the sequence $A'_1, A'_2, \cdots$ consists of elements of $\mathbf{S}$, and it is restrictedly metrically convergent to A by (1.5.δ) and (1.5.ϵ).

Observe that in the above proof we even had $||| A_{\nu'} ||| \leqq ||| A |||$ when A is Hermitian. This could be extended to all A but we shall not pursue this matter further.

§1.6 The desired results are now immediate.

LEMMA 1.6.1. *Restricted metric closure of an algebra* $\mathbf{S}$ *implies its metric closure.*

PROOF. Immediate by Lemma 1.5.4.

THEOREM I. *For an algebra* $\mathbf{S}$ *within a finite* $\mathbf{M}$ *the following eight notions of closure are equivalent to each other:*

α) *Restricted metric closure*
β) *Metric closure*
γ) *Weak (topological) closure*
δ) *Weak convergence closure*
ϵ) *Strong (topological) closure*
θ) *Strong convergence closure*
ζ) *Strongest (topological) closure*
η) *Strongest convergence closure.*

PROOF. The implications $\epsilon) \rightarrow \alpha)$, $\beta) \rightarrow \epsilon)$, $\alpha) \rightarrow \beta)$ were established in Lemmas 1.4.1, 1.4.2, 1.6.1, respectively. Consequently $\alpha) \rightleftarrows \beta) \rightleftarrows \epsilon)$. Also

[11] The set, $| \lambda | \leqq ||| A_{\nu'} |||$, contains the entire spectrum of $A_{\nu'}$.

Lemma 1.3.2 yields $\alpha) \rightleftarrows \theta)$. From [2], p. 396, end of §3, we have $\gamma) \rightleftarrows \epsilon)$. [4], p. 112, beginning of §3, yields $\gamma) \rightleftarrows \zeta)$. Also [4], p. 112, end of §2, implies $\theta) \rightleftarrows \eta)$. These equivalences give together

$$\alpha) \rightleftarrows \beta) \rightleftarrows \gamma) \rightleftarrows \epsilon) \rightleftarrows \theta) \rightleftarrows \zeta) \rightleftarrows \eta).$$

And obviously

$$\gamma) \rightarrow \delta) \rightarrow \theta).$$

Thus δ) also is equivalent to the others.

Thus all those topologizations of the space of all operators, which may be reasonably used in this connection, turn out to be equivalent to each other for the algebras **S** in a factor **M** which is in a finite case.[4]

We also state explicitly

THEOREM II. *The notion of a subring* **S** *of* **M**, *for a finite* **M**, *is purely algebraical.*

PROOF. The equivalences of Theorem I now permit us to apply the argument given at the end of §1.2.

Thus the desideratum expressed at the end of §1.1 is fully realized.

CHAPTER II. THE ALGEBRAICAL TYPE AND ITS OPERATIONS. THE FUNDAMENTAL GROUP

§2.1 We shall call the abstraction of a ring **M** with respect to algebraical isomorphism, its algebraical type, i.e. two rings $\mathbf{M}_1$ and $\mathbf{M}_2$ will be said to have the same algebraical type if and only if they are algebraically isomorphic. We denote the algebraic type of the ring **M** by $\bar{\mathbf{M}}$. Notions in **M** which are purely algebraical, i.e. invariant under algebraical isomorphisms, will be said to be notions concerning the algebraical type $\bar{\mathbf{M}}$ (and not **M** itself).[12]

Thus the properties enumerated in §1.1 (or equally in [9]) are properties of $\bar{\mathbf{M}}$ for every ring **M**. The notion of a subring of **M** concerns $\bar{\mathbf{M}}$ when **M** is a factor in a finite class (cf. Theorem II). This is also true of $D_{\mathbf{M}}(E)$, $Tr_{\mathbf{M}}(A)$, [[A]] (cf. the beginning of §1.2), and the various notions of closure, as enumerated in Theorem I, when they concern an algebra **S**. Thus when **M** is in a finite class, the properties concerning $\bar{\mathbf{M}}$ are particularly extensive.

§2.2 As in the introduction, §2, let for each $i = 1, 2$ a Hilbert space $\mathfrak{H}_i$ and an operator ring $\mathbf{M}_i$ in $\mathfrak{H}_i$ be given. We generalize B), loc. cit., as follows:

B^*) General (algebraic ring) isomorphism of $\mathbf{M}_1$ and $\mathbf{M}_2$. This is a one-to-one mapping of $\mathbf{M}_1$ on $\mathbf{M}_2$ which leaves the entity 1 and the operations A^*, $A + B$ invariant; while it carries the entity αA (α any complex number) into either αA always (case a') or into $\bar{\alpha} A$ always (case b') and AB into AB always

[12] This definition should be compared and contrasted with that of spatial type in §3.3.

(case a'') or into BA always (case b''). More specifically, we talk according to which combination of the above cases occurs, of an $[a', a'']$ (or "proper") $[b', a'']$ (or "conjugate") $[a'b'']$ (or "dual") or $[b', b'']$ (or "conjugate dual") isomorphism.

The existence of a general isomorphism, of any one of the above four kinds, between $\mathbf{M}_1$ and $\mathbf{M}_2$ is clearly a property of the types $\overline{\mathbf{M}}_1$ and $\overline{\mathbf{M}}_2$. If a (proper) isomorphism holds, we have $\overline{\mathbf{M}}_1 = \overline{\mathbf{M}}_2$. If a conjugate or a dual isomorphism holds, then $\overline{\mathbf{M}}_1$ determines $\overline{\mathbf{M}}_2$ uniquely, assuming its existence. Hence we may express these relationships by $\overline{\mathbf{M}}_1^{\backprime} = \overline{\mathbf{M}}_2$ or $\overline{\mathbf{M}}_1' = \overline{\mathbf{M}}_2$ respectively.

We must now deduce the existential and other properties of the operations $\overline{\mathbf{M}}^{\backprime}$ and $\overline{\mathbf{M}}'$.

§2.3 Let a Hilbert space $\mathfrak{H}$ be given. Modify the fundamental operations $\alpha f, f + g, (f, g)$ in $\mathfrak{H}$ by replacing them by $\bar{\alpha} f, f + g, \overline{(f, g)}$. Denote the set $\mathfrak{H}$ with the new definition of its fundamental operations by $\mathfrak{H}_c$. Clearly $\mathfrak{H}_c$ is also a Hilbert space; every (linear bounded) operator A in $\mathfrak{H}$ is also one in $\mathfrak{H}_c$, and every ring of operators $\mathbf{M}$ in $\mathfrak{H}$ is also one in $\mathfrak{H}_c$. But we shall denote the A, $\mathbf{M}$ of $\mathfrak{H}$, when considered in $\mathfrak{H}_c$, by A_c, $\mathbf{M}_c$.

Now consider an operator ring $\mathbf{M}$ in $\mathfrak{H}$. The identical mapping $\mathfrak{J}^\circ$ then maps $\mathbf{M}$ in $\mathfrak{H}$ on $\mathbf{M}_c$ in $\mathfrak{H}_c$ and it is in this aspect a conjugate isomorphism of $\mathbf{M}$ and $\mathbf{M}_c$.

Consider next the mapping $\mathfrak{J}^*$, $A \rightarrow A^*$. This maps $\mathbf{M}$ in $\mathfrak{H}$ on $\mathbf{M}$ in $\mathfrak{H}$ and it is, in this aspect, a conjugate dual isomorphism of $\mathbf{M}$ and $\mathbf{M}$. Consequently $\mathfrak{J}^*\mathfrak{J}^\circ$ (i.e. $\mathfrak{J}^*$ in a new aspect) is a dual isomorphism of $\mathbf{M}$ and $\mathbf{M}_c$.

Thus we have

THEOREM III. *$\overline{\mathbf{M}}^{\backprime}$ and $\overline{\mathbf{M}}'$ always exist and are single-valued. They are both equal to the above $\overline{\mathbf{M}}_c$. We shall denote them by $\overline{\mathbf{M}}^c$.*

Clearly

$$\overline{\mathbf{M}}^{cc} = \overline{\mathbf{M}} \tag{2.3.α}$$

since $\mathfrak{H}_{c,c} = \mathfrak{H}$, $A_{c,c} = A$, $\mathbf{M}_{c,c} = \mathbf{M}$.

Observe that the definitions of a factor and of the case to which it belongs are unaffected by conjugate isomorphisms. For these definitions make use of αA only in defining the set of all $\alpha 1$ but never for a specific numerical value of α. To see this, reconsider [5], p. 138, Def. 3.1.2; p. 173, Theorem IX. Hence the operation $\overline{\mathbf{M}}^c$ (viewed as $\overline{\mathbf{M}}^{\backprime}$) does not affect the factor character of $\mathbf{M}$ nor the case to which it belongs.

§2.4 In what follows we shall consider matrices of operators. Given a $p = 1, 2, \cdots, \infty$ a *p-order matrix* $\langle A_{t,s} \rangle$ is a scheme or array of operators $A_{t,s}$ in which the indices s and t range from 1 to p for a finite p and over $1, 2, \cdots$ if $p = \infty$. (This last will be abbreviated $s, t \leqq p$ in both cases.) An $\langle A_{t,s} \rangle$ is finite if there exists a finite $q = 1, 2, \cdots$ such that $A_{t,s} = 0$ if either $t > q$ or $s > q$. For a finite p this is automatically fulfilled, but it is a real restriction for $p = \infty$.

Now as in the Introduction, §2, let, for each $i = 1, 2$, a Hilbert space $\mathfrak{H}_i$

and an operator ring $\mathbf{M}_i$ in $\mathfrak{H}_i$ be given. We generalize B), loc. cit., as follows:

B_p) *p-matrix* (*algebraic ring*) *isomorphism of* $\mathbf{M}_1$ *over* $\mathbf{M}_2$. This is a one-to-one mapping of $\mathbf{M}_1$ on a set $\mathfrak{J}$ of p-order matrices $\langle A_{t,s}\rangle$ over $\mathbf{M}_2$ (i.e. $A_{t,s} \in \mathbf{M}_2$) which is algebraically ring-isomórphic in the matrix sense, i.e. it carries 1, αA°, $A^{\circ *}$ $A^\circ + B^\circ$, $A^\circ B^\circ$ in $\mathbf{M}_1$ into $\langle \delta_{t,s}1\rangle$, $\langle \alpha A_{t,s}\rangle$ $\langle A^*_{s,t}\rangle$, $\langle A_{t,s} + B_{t,s}\rangle$, $\langle \sum_{r=1}^{p} A_{r,s}B_{t,r}\rangle$ in $\mathfrak{J}$. The $\sum_{r=1}^{p}$ in the above expression, when $p = \infty$ must converge in the sense of the strong operator topology, irrespective of the order of summation.

The set $\mathfrak{J}$ must, at any rate, contain all finite matrices $\langle A_{t,s}\rangle$ over $\mathbf{M}_2$. (Cf. [5], pp. 136–137, Lemma 2.4.3, where the same notions are used, the same multiplication convention obtains, as well as the same notion of convergence for $\sum_{r=1}^{\infty}$.)

For a given p, $\mathbf{M}_1$, $\mathbf{M}_2$ the existence of such a p-matrix isomorphism of $\mathbf{M}_1$ over $\mathbf{M}_2$ is a property of types $\overline{\mathbf{M}}_1$, $\overline{\mathbf{M}}_2$. This is obvious for $p = 1, 2, \cdots$ while for $p = \infty$ it is only necessary to remember the purely algebraical character of the notion of strong convergence (cf. the beginning of §1.1, or [9], Theorem II). Whether $\overline{\mathbf{M}}_1$ (assuming its existence) is uniquely determined by $\overline{\mathbf{M}}_2$ or not, for a given $p = 1, 2, \cdots, \infty$ depends obviously on whether the set $\mathfrak{J}$ in B_p) is uniquely determined or not. For $p = 1, 2, \cdots$ (p finite), the last part of B_p) implies that the set $\mathfrak{J}$ must be the set of all $\langle A_{t,s}\rangle$ with $A_{t,s} \in \mathbf{M}_2$ since they are all finite. Hence in this case $\overline{\mathbf{M}}_1$ is uniquely determined. For $p = \infty$ the question of uniqueness is not so immediately settled. We shall prove that $\mathfrak{J}$ and with it $\overline{\mathbf{M}}_1$ are again unique, but we delay this discussion to the next section. Nevertheless, for $p = 1, 2, \cdots, \infty$ we express the existence of a p-order matrix isomorphism of $\mathbf{M}_1$ over $\mathbf{M}_2$ as a relationship between $\overline{\mathbf{M}}_1$ and $\overline{\mathbf{M}}_2$ by $\overline{\mathbf{M}}_1 = \overline{\mathbf{M}}_2^p$. But then, quite apart from the question of existence, we may assume that $\overline{\mathbf{M}}^p$ is single-valued only for a finite p. At present we must consider the possibility that for $p = \infty$, $\overline{\mathbf{M}}^p$ is many valued.

The general existence of $\overline{\mathbf{M}}^p$ belongs in abstract algebra, but we shall need $\overline{\mathbf{M}}^p$ as an operator ring. For this reason, as well as for the sake of general orientation, we shall construct it explicitly within the framework of [5], pp. 135–137, §2.4.

Let a Hilbert space $\mathfrak{H}$ and a ring $\mathbf{M}$ in it be given, and a $p = 1, 2, \cdots, \infty$. We perform the constructions of [5], loc. cit., the $\mathfrak{H}_2$ there being our $\mathfrak{H}$ and the $\mathfrak{H}_1$ there, any p-dimensional unitary space. (This is not the $\mathfrak{H}_1$ of B_p) above. We conform to the notations of [5], loc. cit.) Form $\mathfrak{H}_1 \otimes \mathfrak{H}_2$ and the ring $\mathbf{B}\otimes$ of all its bounded operators. Then all elements of $\mathbf{B}\otimes$ can be represented as p order matrices $\langle A_{t,s}\rangle$ ($t, s \leqq p$, $A_{t,s}$ in $\mathfrak{H}_2 = \mathfrak{H}$). (Cf. [5], p. 136, Def. 2.4.2 and Lemma 2.4.2.) For the given ring $\mathbf{M}$, form the set $\mathbf{M}_1$ consisting of all operators in $\mathfrak{H}_1 \otimes \mathfrak{H}_2$ which have a matrix $\langle A_{t,s}\rangle$ with all $A_{t,s} \in \mathbf{M}$. $\mathfrak{J}$ is the set of all those $\langle A_{t,s}\rangle$ with $A_{t,s} \in \mathbf{M}$ which correspond to some operator in $\mathfrak{H}_1 \otimes \mathfrak{H}_2$. From this it follows immediately that $\mathfrak{J}$ contains all finite matrices $\langle A_{t,s}\rangle$ with $A_{t,s} \in \mathbf{M}$.[13] Now $\mathbf{M}_1$ in $\mathfrak{H}_1 \otimes \mathfrak{H}_2$ and $\mathbf{M}$ in $\mathfrak{H} = \mathfrak{H}_2$ together with

[13] For a finite matrix $\langle A_{t,s}\rangle$ an operator A° in $\mathfrak{H}_1 \otimes \mathfrak{H}_2$ with $A^\circ \backsim \langle A_{t,s}\rangle$ can be constructed in the sense of [5], p. 136, Def. 2.4.2, without any convergence difficulties.

the interpretation of the matrix scheme $\langle A_{t,s}\rangle$ which we are now using, clearly fulfill all the requirements of B_p). (We have our $\mathfrak{H}_1 \otimes \mathfrak{H}_2$, $\mathfrak{H} = \mathfrak{H}_2$, $\mathbf{M}_1$, $\mathbf{M}$ in place of the $\mathfrak{H}_1$, $\mathfrak{H}_2$, $\mathbf{M}_1$, $\mathbf{M}_2$ of B_p).)

Summing up the results obtained so far, we have

THEOREM IV. *$\overline{\mathbf{M}}^p$ exists for all $p = 1, 2, \cdots, \infty$ and it is certainly one-valued for $p = 1, 2, \cdots$ (p finite).*

Our definitions in B_p) and that of §2.3 give together

$$(\overline{\mathbf{M}}^c)^p = (\overline{\mathbf{M}}^p)^c, \qquad (p = 1, 2, \cdots), \tag{2.4.α}$$

and B_p) gives

$$(\overline{\mathbf{M}}^p)^q = \overline{\mathbf{M}}^{pq} \qquad (p, q = 1, 2, \cdots). \tag{2.4.β}$$

(The exponent ∞ will be considered in the next section.)

§2.5 In this section we shall make use of the operations $A_{(\mathfrak{M})}$, $\mathbf{M}_{(\mathfrak{M})}$ defined in [1], p. 78, and [5], p. 186, Def. 11.3.1. If A is an operator in $\mathfrak{H}$ which is reduced by the closed linear set $\mathfrak{M}$ then $A_{(\mathfrak{M})}$ is its part in $\mathfrak{M}$. And $\mathbf{M}_{(\mathfrak{M})}$ is the set of all $A_{(\mathfrak{M})}$ for all $A \in \mathbf{M}$ which are reduced by $\mathfrak{M}$. $A_{(\mathfrak{M})}$, $\mathbf{M}_{(\mathfrak{M})}$ are in $\mathfrak{M}$ while A, $\mathbf{M}$ were in $\mathfrak{H}$.

LEMMA 2.5.1. *Assume $E = P_{\mathfrak{M}} \in \mathbf{M}$. Let $\mathbf{M}^E$ be the set of all $A \in \mathbf{M}$ for which $EA = AE = A$. Then we have*

(i) *$A_{(\mathfrak{M})}$ can be formed for all $A \in \mathbf{M}^E$.*

(ii) *$\mathbf{M}_{(\mathfrak{M})}$ coincides with $(\mathbf{M}^E)_{(\mathfrak{M})}$.*

(iii) *$A \rightarrow A_{(\mathfrak{M})}$ is a one-to-one mapping of $\mathbf{M}^E$ on $\mathbf{M}_{(\mathfrak{M})}$ which carries E (in $\mathbf{M}^E$, i.e. in $\mathfrak{H}$) into 1 (in $\mathbf{M}_{(\mathfrak{M})}$ i.e. in $\mathfrak{M}$) and leaves the operations αA, A^*, $A + B$, AB invariant.*

PROOF. Ad (i) Every $A \in \mathbf{M}^E$ commutes with $E = P_{(\mathfrak{M})}$, i.e. it is reduced by $\mathfrak{M}$.

Ad (ii) $A \in \mathbf{M}$ implies $EAE \in \mathbf{M}$ and thence clearly $EAE \in \mathbf{M}^E$. Now if $\mathfrak{M}$ reduces A, then both A and EAE have the same part in $\mathfrak{M}$, i.e. $A_{(\mathfrak{M})} = (EAE)_{(\mathfrak{M})}$. Thus $(\mathbf{M}^E)_{(\mathfrak{M})} \supseteq \mathbf{M}_{(\mathfrak{M})}$. But $\mathbf{M}^E \subseteq \mathbf{M}$ hence also $(\mathbf{M}^E)_{(\mathfrak{M})} \subset \mathbf{M}_{(\mathfrak{M})}$. So $(\mathbf{M}^E)_{(\mathfrak{M})} = \mathbf{M}_{(\mathfrak{M})}$.

Ad (iii) Every $A \in \mathbf{M}^E$ is reduced by $\mathfrak{M}$ (cf. (i) above) and its part in $\mathfrak{M}^\perp$ ($= \mathfrak{H} \ominus \mathfrak{M}$) is clearly 0. So we may view it as an operator in $\mathfrak{M}$ as well as in $\mathfrak{H}$, and in its former aspect it is $A_{(\mathfrak{M})}$. All our assertions are now immediate.

Consider now $\mathbf{M}_1$, $\mathbf{M}_2$ and $\mathfrak{H}_1$, $\mathfrak{H}_2$ and an ∞-matrix isomorphism of $\mathbf{M}_1$ over $\mathbf{M}_2$ as in B_∞) in the last section. For any $q = 1, 2, \cdots$ form the set $\mathfrak{J}^q$ of all matrices $\langle A_{t,s}\rangle$ for which all $A_{t,s} \in \mathbf{M}_2$ and where $A_{t,s} = 0$ when $t > q$ or $s > q$. Form also the matrix $\langle e^q_{t,s}1\rangle$ where $e^q_{t,s} = 1$ if $t = s \leqq q$ and $e^q_{t,s} = 0$ otherwise. Then it is easy to establish that

LEMMA 2.5.2. (i) $\mathfrak{J}^q \subseteq \mathfrak{J}$.

(ii) *$\langle e^q_{t,s}\rangle \in \mathfrak{J}^q$ corresponds to a projection $E^q \in \mathbf{M}$. Put $E^q = P_{\mathfrak{M}^q}$.*

(iii) *$E^1 \leqq E^2 \leqq \cdots$ and $\lim_{q\to\infty} E^q = 1$ in the strong sense.*

(iv) *Every matrix $\langle A_{t,s}\rangle \in \mathfrak{J}^q$ can be viewed both as a q^{th} order matrix and as an ∞-order matrix. It corresponds in its former aspect to an element of a ring*

$\mathbf{M}_2^q$ *with* $\overline{\mathbf{M}_2^q} = \overline{\mathbf{M}}_2^q$ *and in its latter aspect to an element of* $(\mathbf{M}_1)^{E^q} \subseteq \mathbf{M}_1$. *The latter correspondence can be continued by Lemma* 2.5.1 *to one with an element of* $(\mathbf{M}_1)_{(\mathfrak{M}^q)}$. *These elements exhaust the sets* $\mathbf{M}_2^q$, $\mathbf{M}_1^{E^q}$ *and* $(\mathbf{M}_1)_{(\mathfrak{M}^q)}$, *respectively.*

(v) *The correspondence of (iv) is an algebraic ring isomorphism of* $\mathbf{M}_2^q$ *and* $(\mathbf{M}_1)_{(\mathfrak{M}^q)}$. *Hence* $\overline{(\mathbf{M}_1)_{(\mathfrak{M}^q)}} = \overline{\mathbf{M}_2^q} = \overline{\mathbf{M}}_2^q$.

PROOF. Ad (i). Immediate by the final clause of B_∞).

Ad (ii). Clearly $\langle e_{t,s}^q \cdot 1\rangle \in \mathfrak{J}^q$ hence it is $\in \mathfrak{J}$ and thus corresponds to an element of $\mathbf{M}_1$. The rules of matrix computation show that this element is a projection.

Ad (iii). The rules of matrix computation show that $E^1 \leqq E^2 \leqq \cdots$. Now if $f = \langle f_1, f_2, \cdots\rangle \in \mathfrak{H}_1 \otimes \mathfrak{H}_2$ then $E^q f = \langle f_1, \cdots, f_q, 0, \cdots\rangle \to f$ as $q \to \infty$. Hence $\lim_{q\to\infty} E^q$ is .

Ad (iv). The $\langle A_{t,s}\rangle \in \mathfrak{J}^q$ are clearly characterized by $\langle e_{t,s}^q \cdot 1\rangle \cdot \langle A_{t,s}\rangle = \langle A_{t,s}\rangle\langle e_{t,s}^q \cdot 1\rangle = \langle A_{t,s}\rangle$ and thus we may add by (i), $\langle A_{t,s}\rangle \in \mathfrak{J}$. Hence the second correspondence mentioned maps the $\langle A_{t,s}\rangle \in \mathfrak{J}^q$ precisely on the $A^\circ \in \mathbf{M}$ with $E^q A^\circ = A^\circ E^q = A^\circ$ i.e. on the $A^\circ \in (\mathbf{M}_1)^{E^q}$. All other assertions are obvious, remembering in particular Lemma 2.5.1.

Ad (v). The corresponding assertions for the relationship between $\mathbf{M}_2^q$ and $(\mathbf{M}_1)^{E^q}$ follow immediately from the rules of matrix computation. They are transferred to $\mathbf{M}_2^q$ and $(\mathbf{M}_1)_{(\mathfrak{M}_q)}$ by using Lemma 2.5.1.

We now go further:

LEMMA 2.5.3. *A matrix* $\langle A_{t,s}\rangle \in \mathfrak{J}^q$ *corresponds to three operators in the sense of (iv) in Lemma* 2.5.2, *i.e. to an operator of* $\mathbf{M}_2^q$ *to one in* $(\mathbf{M}_1)^{E^q}$ *and to one in* $(\mathbf{M}_1)_{(\mathfrak{M}^q)}$. *All three operators have the same bound,* $|||\cdots|||$ *and we denote this common bound by* $|||\langle A_{t,s}\rangle|||$.

PROOF. The operator in $(\mathbf{M}_1)^{E^q}$ and that one in $(\mathbf{M}_1)_{(\mathfrak{M}^q)}$ have the same bound, since we are considering the same operator, once in $\mathfrak{H}_1$ and once in $\mathfrak{M}^q$.[14] The operator in $\mathbf{M}_2^q$ and that one in $(\mathbf{M}_1)_{(\mathfrak{M}^q)}$ have the same bound since they obtain from one another by an algebraical ring isomorphism (cf. (v) in Lemma 2.5.2) and since the numerical value of the bound is a purely algebraical notion (cf. §1.1.). This completes the proof.

LEMMA 2.5.4. *Consider a matrix* $\langle A_{t,s}\rangle$ *all* $A_{t,s} \in \mathbf{M}_2$. *Define the matrices* $\langle A_{t,s}^q\rangle \in \mathfrak{J}^q$ *by the equations* $A_{t,s}^q = A_{t,s}$, *if* $t \leqq q$ *and* $s \leqq q$ *and* $A_{t,s}^q = 0$ *otherwise. Then* $\langle A_{t,s}\rangle \in \mathfrak{J}$ *if and only if the numerical sequence* $|||\langle A_{t,s}^1\rangle|||$, $|||\langle A_{t,s}^2\rangle|||, \cdots$ *(cf. Lemma* 2.5.3 *above) is bounded.*

PROOF. Necessity: Assume $\langle A_{t,s}\rangle \in \mathfrak{J}$ and let it correspond to $A^\circ \in \mathbf{M}_1$. Then, clearly $\langle A_{t,s}^q\rangle = \langle e_{t,s}^q \cdot 1\rangle \cdot \langle A_{t,s}\rangle\langle e_{t,s}^q \cdot 1\rangle$ i.e. $\langle A_{t,s}^q\rangle$ corresponds to the operator $E^q A^\circ E^q \in \mathbf{M}_1$. Clearly $E^q A^\circ E^q \in (\mathbf{M}_1)^{E^q}$. Thus the second definition of Lemma 2.5.3 gives $|||\langle A_{t,s}^q\rangle||| = |||E^q A^\circ E^q|||$. Now we have in $\mathfrak{H}_1$ that $|||E^q A^\circ E^q||| \leqq |||E^q||| \cdot |||A^\circ||| \cdot |||E^q||| = |||A^\circ|||$. So $|||\langle A_{t,s}^q\rangle||| \leqq |||A^\circ|||$. This holds for $q = 1, 2, \cdots$ and hence the sequence $|||\langle A_{t,s}^1\rangle|||$, $|||\langle A_{t,s}^2\rangle|||, \cdots$ is bounded.

Sufficiency: Assume that the numerical sequence $|||\langle A_{t,s}^1\rangle|||$, $|||\langle A_{t,s}^2\rangle|||, \cdots$

---

[14] The part of the first-mentioned operator in $(\mathfrak{M}^q)^\perp$ is 0.

is bounded. $\langle A^q_{t,s}\rangle \in \mathfrak{J}^q$ so it corresponds to an $A^q \in (\mathbf{M}_1)^{E^q} \subsetneqq \mathbf{M}_1$. Owing to the definition of Lemma 2.5.3, $|||\, \langle A^q_{t,s}\rangle \,||| = |||\, A^q \,|||$ hence

(2.5.α) The (numerical) sequence $|||\, A^1 \,|||, |||\, A^2 \,|||, \cdots$ is bounded.

For $q \geqq r$ clearly $\langle e^r_{t,s}\cdot 1\rangle\langle A^q_{t,s}\rangle\langle e^r_{t,s}\cdot 1\rangle = \langle A^r_{t,s}\rangle$ hence $E^rA^qE^r = A^r$.

Now consider an $f \in \mathfrak{M}^r$.

For $q \geqq r$, let $f_q = A^qf$. Since $\|f_q\|^2 \leqq |||\, A^q \,|||^2 \cdot \|f\|^2$, these f_q's are bounded. Also $f \in \mathfrak{M}^r \subset \mathfrak{M}^q$ so $E^qf = f$. Hence for $q' \geqq q \geqq r$, $E^qf_{q'} = E^qA^{q'}f = E^qA^{q'}E^qf = A^qf = f_q$. So $f_{q'} - f_q$ and f_q are orthogonal, $\|f_{q'} - f_q\|^2 + \|f_q\|^2 = \|f_{q'}\|^2$ i.e.

$$\|f_{q'} - f_q\|^2 = \|f_{q'}\|^2 - \|f_q\|^2 \qquad \text{for } q' \geqq q \geqq r \tag{2.5.β}$$

Thus the $\|f_r\|^2, \|f_{r+1}\|^2, \cdots$ form a monotone bounded sequence, and hence the $\lim_{q\to\infty} \|f_q\|^2$ exists. Therefore the right-hand side of (2.5.β) converges to zero for $q, q' \to \infty$. Hence (2.5.β) now implies that strong limit A^qf exists for $q \to \infty$. The restriction $q \geqq r$ may now be dropped.

The set of all f for which A^qf is strongly convergent is closed, since the $|||\, A^q \,|||$ are uniformly bounded. This set is obviously linear and by the above it contains $\mathfrak{M}^1, \mathfrak{M}^2, \cdots$. Thus we see

(2.5.γ) $\begin{cases}\text{The sequence } A^1, A^2, \cdots \text{ converges for all } f \text{ in the}\\ \text{closed linear set determined by } \mathfrak{M}^1, \mathfrak{M}^2, \cdots.\end{cases}$

Since the $\mathfrak{M}^1, \mathfrak{M}^2, \cdots$ are together dense, the set of (2.5.γ) is necessarily $\mathfrak{H}$ itself. Denote the limit of $A^1f, A^2f, \cdots$ by Af. Then A belongs to $\mathbf{M}$ along with $A^1, A^2, \cdots$. Thus we have shown

(2.5.δ) $\begin{cases}\text{The sequence } A^1, A^2, \cdots \text{ converges for all } f \text{ in } \infty \otimes \mathfrak{H}\\ \text{to an } A \in \mathbf{M} \text{ in the strong sense.}\end{cases}$

Let A correspond to the matrix $\langle B_{t,s}\rangle$ in B_∞). Of course $\langle B_{t,s}\rangle \in \mathfrak{J}$. Since for $q \geqq r$, $E^rA^qE^r = A^r$, (2.5.δ) implies that $E^rAE^r = A^r$. Hence $\langle e^r_{t,s}\cdot 1\rangle\langle B_{t,s}\rangle\langle e^r_{t,s}\cdot 1\rangle = \langle A^r_{t,s}\rangle$. Thus the rules of matrix computation yield $B^r_{t,s} = A^r_{t,s}$. Now let t, s be given and choose $r = \text{Max}\,(t, s)$. Then the above formula gives $B_{t,s} = A_{t,s}$. Since t, s were arbitrary, this means $\langle B_{t,s}\rangle = \langle A_{t,s}\rangle$ and consequently $\langle A_{t,s}\rangle \in \mathfrak{J}$ too.

Thus the proof is completed.

LEMMA 2.5.5. *The set $\mathfrak{J}$ can be uniquely and purely algebraically characterized in terms of $\mathbf{M}_2$ alone.*

PROOF. A unique characterization of $\mathfrak{J}$ was given in Lemma 2.5.4. By using the first definition of Lemma 2.5.3 for $|||\, \langle A^q_{t,s}\rangle \,|||$ that one in terms of $\mathbf{M}^q_2$ it becomes clear that this characterization is purely algebraical in terms of $\mathbf{M}_2$ and of the $\mathbf{M}^q_2$, $q = 1, 2, \cdots$. But we have already reduced the $\mathbf{M}^q_2$ i.e. the $\mathbf{M}^q_2 = \mathbf{M}^q_2$ for q finite, to $\mathbf{M}_2$ by Theorem IV.

The proof is thus completed.

The remarks made in §2.4, preceding Theorem IV, in conjunction with the above Lemma 2.5.5, permit us now to assert

THEOREM IV′. *$\overline{\mathbf{M}}^{\infty}$ too is one-valued.*

Our definitions in B_∞ and that of §2.3 give together

$$(\overline{\mathbf{M}}^{e})^{\infty} = (\overline{\mathbf{M}}^{\infty})^{e}. \tag{2.5.ϵ}$$

(This corresponds to (2.4.α). We could extend (2.4.β) too, but we shall not need it here. Furthermore this extension is an easy consequence of Lemma 3.1.6.)

§2.6 We undertake now to find the inverse operation to $\mathbf{N}^p$. At first we restrict neither the ring $\mathbf{N}$ nor the exponent $p = 1, 2, \cdots, \infty$.

DEFINITION 2.6.1. *A system of p^2 operators $W_{\mu,\nu}$ ($\mu, \nu \leqq p$) is a system of p-matrix units if it possesses the following properties:*

(i) $W^*_{\mu,\nu} = W_{\nu,\mu}$

(ii) $W_{\mu,\nu}W_{\sigma,\omega} = W_{\mu,\omega}$ *for* $\nu = \sigma$

$= 0$ *if* $\nu \neq \sigma$.

Some immediate properties of these systems are given in

LEMMA 2.6.1. *Every system of p-matrix units $W_{\mu,\nu}$ ($\mu, \nu \leqq p$) possesses the following properties:*

(i) *The $W_{\mu,\mu}$ ($\mu \leqq p$) are pairwise orthogonal projections.*

(ii) *Every $W_{\mu,\nu}$ is partially isometric with the initial projection $W_{\nu,\nu}$ and the final projection $W_{\mu,\mu}$* (cf. [5], §4.3).

(iii) *Form $E_0 = \sum_{\mu=1}^{p} W_{\mu,\mu}$. This $\sum_{\mu=1}^{p}$ is either a finite sum or when $p = \infty$ converges in the sense of the strong operator topology, irrespective of the order in which the $\mu = 1, 2, \cdots$ are gone through. This E_0 is a projection and it is the unit of the system $W_{\mu,\nu}$ i.e. always*

$$E_0W_{\mu,\nu} = W_{\mu,\nu}E_0 = W_{\mu,\nu}.$$

PROOF. Ad (i), (iii): The convergence in (iii) is a consequence of (i), as is seen from familiar considerations concerning pairwise orthogonal projections (cf. e.g. [5], pp. 76–78, or the proof of (2.5.δ) in the proof of Lemma 2.5.4 above). All other assertions are immediately verified by using (i), (ii) in Def. 2.6.1.

Ad (ii). Definition 2.6.1 gives directly

$$W^*_{\mu,\nu}W_{\mu,\nu} = W_{\nu,\nu}, \qquad W_{\mu,\nu}W^*_{\mu,\nu} = W_{\mu,\mu}.$$

Since $W_{\nu,\nu}$, $W_{\mu,\mu}$ are projections by (i) above, these formulae imply our assertions by [5], p. 142, Lemma 4.3.2 (the last two criteria), and p. 141, Lemma 4.3.1.

We prove now

THEOREM V. *$\overline{\mathbf{M}} = \overline{\mathbf{N}}^p$ is equivalent to this: There exists a system of p-matrix units $W_{\mu,\nu} \in \mathbf{M}$ ($\mu, \nu \leqq p$) with the unit* 1 (*cf.* (iii) *in Lemma* 2.6.1) *and with the following further property:*

$W_{1,1}$ is a projection. Let $\mathfrak{M}$ be the closed linear set with $P_{\mathfrak{M}} = W_{1,1}$. Then $\mathbf{M}_{(\mathfrak{M})}$ must be algebraically isomorphic to $\mathbf{N}$.[15]

(Concerning the omission of this extra condition, cf. the corollary below.)

PROOF. Sufficiency: Assume the existence of $W_{\mu,\nu}$ $(\mu, \nu \leqq p)$ as described. $W_{1,1} \epsilon \mathbf{M}$ is a projection. Let $\mathfrak{M}$ be such that $P_{\mathfrak{M}} = E = W_{1,1}$.

For every $A^0 \epsilon \mathbf{M}$ define

$$A'_{t,s} = W_{1,s}A^0W_{t,1} \qquad (t, s \leqq p). \tag{2.6.α}$$

Clearly $A'_{t,s} \epsilon \mathbf{M}$ and one verifies immediately that $EA'_{t,s} = A'_{t,s}E = A'_{t,s}$ $(E = W_{1,1})$. Hence $A'_{t,s} \epsilon \mathbf{M}^E$. Now apply Lemma 2.5.1 and form

$$A_{t,s} = (A'_{t,s})_{\mathfrak{M}}. \tag{2.6.β}$$

Accordingly $A_{t,s} \epsilon \mathbf{M}_{(\mathfrak{M})}$.

Thus to every $A^0 \epsilon \mathbf{M}$ there corresponds a p-order matrix $\langle A_{t,s}\rangle$, $(t, s \leqq p)$ with all $A_{t,s} \epsilon \mathbf{M}_{(\mathfrak{M})}$. Let $\mathfrak{J}$ be the set of these $\langle A_{t,s}\rangle$ which correspond to an $A^0 \epsilon \mathbf{M}$.

We shall now show that this correspondence establishes a p-matrix automorphism of $\mathbf{M}$ over $\mathbf{M}_{(\mathfrak{M})}$ in the sense of B_p) in §2,4. That means $\bar{\mathbf{M}} = \bar{\mathbf{M}}^p_{(\mathfrak{M})}$ and as $\overline{\mathbf{M}_{(\mathfrak{M})}} = \bar{\mathbf{N}}$ so $\bar{\mathbf{M}} = \bar{\mathbf{N}}^p$. This will complete the proof of the sufficiency.

We prove the above assertions in three successive steps.

The correspondence is one-to-one—i.e. if A^0, $B^0 \epsilon \mathbf{M}$ have $\langle A_{t,s}\rangle = \langle B_{t,s}\rangle$ then $A^0 = B^0$. Now $\langle A_{t,s}\rangle = \langle B_{t,s}\rangle$ means $A_{t,s} = B_{t,s}$ for t, $s \leqq p$ and hence $(A'_{t,s})_{(\mathfrak{M})} = (B'_{t,s})_{(\mathfrak{M})}$ and $A'_{t,s} = B'_{t,s}$ since both are ϵ $\mathbf{M}^E$. Thus $W_{1,s}A^0W_{t,1} = W_{1,s}B^0W_{t,1}$. Multiplication by $W_{s,1}$ on the left and by $W_{1,t}$ on the right, gives $W_{s,s}A^0W_{t,t} = W_{s,s}B^0W_{t,t}$. If we sum over s and t we obtain $A^0 = B^0$ as desired, since by hypothesis $E = 1 = \sum_{s=1}^{p} W_{s,s}$.

The correspondence is a matrix isomorphism. This can be verified by the use of Def. 2.6.1, $E_0 = \sum_{\mu=1}^{p} W_{\mu,\mu} = 1$ and the equations (2.6.α) and (2.6.β). For these show that by the use of Equation (2.6.α) we correspond to 1, αA^0, A^{0*}, $A^0 + B^0$, A^0B^0 (in $\mathbf{M}$) matrices $\langle\delta_{t,s}E\rangle$, $\langle\alpha A'_{t,s}\rangle$, $\langle A'^*_{s,t}\rangle$, $\langle A'_{t,s} + B'_{t,s}\rangle$ $\langle\sum_{r=1}^{p} A'_{r,s}B'_{t,r}\rangle$ of elements of $\mathbf{M}^E$ and then by the use of Equation (2.6.β) we correspond these to $\langle\delta_{t,s}1\rangle$, $\langle\alpha A_{t,s}\rangle$, $\langle A^*_{s,t}\rangle$, $\langle A_{t,s} + B_{t,s}\rangle$, $\langle\sum_{r=1}^{p} A_{r,s}B_{t,r}\rangle$ whose elements are in $\mathbf{M}_{(\mathfrak{M})}$.

$\mathfrak{J}$ contains all finite matrices. Consider a finite matrix $\langle A_{t,s}\rangle$ where $A_{t,s} \epsilon \mathbf{M}_{(\mathfrak{M})}$. There is a $q < \infty$ such that $A_{t,s} = 0$ when $t > q$ or when $s > q$. Let $A'_{t,s} \epsilon \mathbf{M}^E$ satisfy (2.6.β) and put

$$A^0 = \sum_{t=1}^{q} \sum_{s=1}^{q} W_{s,1}A'_{t,s}W_{1,t}.$$

[15] Observe that in this presentation $\bar{\mathbf{N}}$ is determined by $\bar{\mathbf{M}}$ and $W_{1,1}$ together. Whether $\bar{\mathbf{M}}$ alone suffices to determine $\bar{\mathbf{N}}$ is a different question. If, however, p is finite and $\mathbf{M}$ is a factor, we do have that $\bar{\mathbf{M}}$ and p determine $\bar{\mathbf{N}}$ uniquely if there is such an $\bar{\mathbf{N}}$. For if $\mathfrak{M}^{(1)}$ is the range of $W_{1,1}$ for one choice of the p-matrix units and $\mathfrak{M}^{(2)}$ that for another, then $D_{\mathbf{M}}(\mathfrak{M}^{(1)}) = (1/p)D_{\mathbf{M}}(\mathfrak{H}) = D_{\mathbf{M}}(\mathfrak{M}^{(2)})$. Hence Lemma 2.8.1 (of our present text) implies, $\mathbf{M}_{(\mathfrak{M}^{(1)})} = \mathbf{M}_{(\mathfrak{M}^{(2)})}$. Hence $\overline{\mathbf{M}_{(\mathfrak{M})}}$ depends only on $\bar{\mathbf{M}}$ and p.

If however p is ∞ then $\mathbf{M}$ itself is in an infinite case by (iii) of Lemma 2.7.1. Under these circumstqnces, $\bar{\mathbf{M}}$ fails to determine $\bar{\mathbf{N}}$ as may be seen from δ) in Lemma 3.1.5 (write $\bar{\mathbf{M}}$ for its $\bar{\mathbf{M}}^\infty$).

Then $A^0 \in \mathbf{M}$ and it follows from the rules of Def. 2.6.1 that (2.6.α) holds. Thus $\langle A_{t,s}\rangle$ corresponds to this A^0 and therefore it belongs to $\mathfrak{J}$.

Necessity: Assume $\bar{\mathbf{M}} = \bar{\mathbf{N}}^p$ in the sense of B_p) in §2.4. For every pair $\mu, \nu \leqq p$ form the matrix, $\langle \delta_{t,\nu}\delta_{s,\mu}1\rangle$. This matrix is clearly finite and hence it belongs to $\mathfrak{J}$. Thus it corresponds to an element of $\mathbf{M}$ which we denote by $W_{\mu,\nu}$. Then the matrix computation rules of B_p) give immediately the rules of Def. 2.6.1. Thus the results of Lemma 2.6.1 are now available. Now in the notation of (ii) in Lemma 2.5.2, $E^q = \sum_{\mu=1}^{q} W_{\mu,\mu}$ $(q = 1, 2, \cdots)$. Hence Lemma 2.5.2 (iii) and Lemma 2.6.1 (iii) imply $E_0 = \sum_{\mu=1}^{p} W_{\mu,\mu} = 1$.

Put $P_{\mathfrak{M}} = E = W_{1,1}$. $A^0 \in \mathbf{M}^E$ means $EA^0 = A^0E = A^0$ or, by the above-mentioned rules, that the matrix A^0 has the form $\langle \delta_{t,1}\delta_{s,1}A\rangle$, $A \in \mathbf{N}$. This correspondence between the $A^0 \in \mathbf{M}^E$ and the $A \in \mathbf{N}$ is clearly one-to-one, it carries $E \in \mathbf{M}^E$ (which has the matrix $\langle \delta_{t,1}\delta_{s,1}1\rangle$) into $1 \in \mathbf{N}$ and it leaves the operations αA, A^*, $A + B$, AB invariant. These results, together with Lemma 2.5.1, show therefore that $\mathbf{M}_{(\mathfrak{M})}$ is algebraically isomorphic to $\mathbf{N}$.

Thus all our requirements are fulfilled.

Corollary. *Given* $\mathbf{M}$ *and* p, *the equation* $\bar{\mathbf{M}} = \bar{\mathbf{N}}^p$ *possesses a solution* $\bar{\mathbf{N}}$ *if and only if there exists a system of p-matrix units,* $W_{\mu,\nu} \in \mathbf{M}$ $(\mu, \nu \leqq p)$ *with the unit* 1.

Proof. This becomes clear when we omit the last condition in the above theorem, since the only effect of that condition is to determine $\bar{\mathbf{N}}$ in terms of the $W_{\mu,\nu}$.

Under certain conditions we may even go further.

Lemma 2.6.2. *If* $\mathbf{M}$ *is a factor not in case* (I_n), *and* $p = 2, 3, \cdots$ *is finite, then there is a system of p-matrix units* $W_{\mu,\nu}$, $(\mu, \nu \leqq p)$ *with unit* $E_0 = 1$, $W_{\mu,\nu} \in \mathbf{M}$.

Proof. We first obtain a set $E_1, \cdots, E_p$ of projections in $\mathbf{M}$ such that $D_{\mathbf{M}}(E_i) = D_{\mathbf{M}}(E_1)$ and $\sum_{n=1}^{p} E_n = 1$. Suppose first that $\mathbf{M}$ is in a case (II_1). We may assume that $D_{\mathbf{M}}(1) = 1$. The range of $D_{\mathbf{M}}$ contains $1/p$ (cf. [5], Theorem VIII, p. 172), and hence there is an $E_1 \in \mathbf{M}$ with $D_{\mathbf{M}}(E_1) = 1/p$. Since $D_{\mathbf{M}}(E_1) = 1/p \leqq 1 - 1/p = D_{\mathbf{M}}(1 - E_1)$, there exists a projection $E_2 \leqq 1 - E_1$ with $D_{\mathbf{M}}(E_2) = 1/p = D_{\mathbf{M}}(E_1)$. (Cf. [5], Lemma 8.3.1, p. 167.) Furthermore $D_{\mathbf{M}}(1 - E_1 - E_2) = 1 - 2/p$. If $p > 2$ we can find an $E_3 \leqq 1 - E_1 - E_2$ such that $D_{\mathbf{M}}(E_3) = 1/p$. Thus we may continue until we obtain a set of p pairwise orthogonal projections $E_1, \cdots, E_p \in \mathbf{M}$ with $\sum_{n=1}^{p} E_n = 1$, $D_{\mathbf{M}}(E_n) = 1/p$. If $\mathbf{M}$ is in an infinite case, we can easily modify the proof of [5], Lemma 7.2.3, p. 157, to yield that there are p pairwise orthogonal, dimensionally equivalent, closed, linear sets $\mathfrak{M}_1, \cdots, \mathfrak{M}_p$ which together span $\mathfrak{H}$. In that proof it is only necessary to separate the $\mathbf{R}_i$'s into p sets rather than 2. Then if E_i is the projection on $\mathfrak{M}_i$, $\sum_{n=1}^{p} E_n = 1$ and $D_{\mathbf{M}}(E_n) = \infty = D_{\mathbf{M}}(E_1)$.

Since $D_{\mathbf{M}}(E_t) = D_{\mathbf{M}}(E_1)$ there is a partially isometric $W_{t,1} \in \mathbf{M}$ with initial set $\mathfrak{M}_1$ and final set $\mathfrak{M}_t$ by the definition of the dimensionality function. If we now define $W_{t,s}$ by the equation $W_{t,s} = W_{t,1}W_{s,1}^*$ the properties of the partially isometric $W_{t,1}$ (cf. [5], §4.3) readily yield the equations of Def. 2.6.1 above. Hence the $W_{t,s}$ constitute a set of p matrix units with $E_0 = 1$, $W_{t,s} \in \mathbf{M}$.

LEMMA 2.6.3. *If* **M** *is in case* (I_n) *then* **M** *possesses a set of p-matrix units with* $E_0 = 1$ *if and only if p divides n.*

PROOF. Necessity: If **M** has a set of p-matrix units, with $E_0 = 1$, then by Lemma 2.6.1, the $W_{1,1}, \cdots, W_{p,p}$ are p pairwise orthogonal projections each with same relative dimension, and whose sum is 1. It follows that $D_{\mathbf{M}}(W_{t,t}) = (1/p)D_{\mathbf{M}}(\mathfrak{H})$. If we take $D_{\mathbf{M}}(\mathfrak{H}) = n$, we know that $D_{\mathbf{M}}(W_{t,t}) = (1/p)D_{\mathbf{M}}(\mathfrak{H})$ is a whole number and hence p divide n. (Cf. [5], Theorem VIII, p. 172.)

Sufficiency: If p divides n, we can find p dimensionally equivalent pairwise orthogonal projections $E_1, \cdots, E_p \in \mathbf{M}$ and with $\sum_{n=1}^{p} E_n = 1$. For if $D_{\mathbf{M}}(\mathfrak{H}) = n$, we can find a projection $E_1 \in \mathbf{M}$ with $D_{\mathbf{M}}(E_1) = n/p$ and we proceed as in the proof of Lemma 2.6.2 to find $E_2, \cdots, E_p$. The rest of that proof then applies to the present situation.

Theorem V yields

LEMMA 2.6.4. (i) *If* **M** *is a factor, not in case* (I_n) *and p is a finite integer, then there is an* **N** *such that* $\bar{\mathbf{N}}^p = \bar{\mathbf{M}}$.

(ii) *If* **M** *is in case* (I_n) *there is an* **N** *such that* $\bar{\mathbf{N}}^p = \bar{\mathbf{M}}$ *if and only if p divides n.*

(iii) *If there is an* **N** *such that* $\bar{\mathbf{N}}^p = \bar{\mathbf{M}}$ *for* $p < \infty$, $p = 2, 3, \cdots$, *then there is a manifold* $\mathfrak{M}$ *such that* $\overline{\mathbf{M}_{(\mathfrak{M})}} = \bar{\mathbf{N}}$ *and* $D_{\mathbf{M}}(\mathfrak{M}) = (1/p)D_{\mathbf{M}}(\mathfrak{H})$.

As we remarked in footnote 15, we may add

LEMMA 2.6.5. *If p is finite,* **M** *is a factor and if* **N** *is such that* $\bar{\mathbf{N}}^p = \bar{\mathbf{M}}$ *then* $\bar{\mathbf{N}}$ *is uniquely determined.*

§2.7 It remains to consider the nature of **N** and also what happens if $p = \infty$.

LEMMA 2.7.1. *We have for any fixed* $p = 1, 2, \cdots, \infty$,

(i) $\bar{\mathbf{N}}$ *is a factor if and only if* $\bar{\mathbf{N}}^p$ *is one.*

(ii) *If* $\bar{\mathbf{N}}, \bar{\mathbf{N}}^p$ *are factors, then they belong to the same one of the Cases* (I), (II), (III).

(iii) *If* $\bar{\mathbf{N}}, \bar{\mathbf{N}}^p$, *are factors, then* $\bar{\mathbf{N}}^p$ *is in a finite case if and only if* $\bar{\mathbf{N}}$ *is in a finite case and p is finite.*

PROOF. Choose an **M** with $\bar{\mathbf{M}} = \bar{\mathbf{N}}^p$.

Ad (i). An $A^0 \in \mathbf{M}$ belongs to the center of **M** if and only if its matrix $\langle A_{t,s}\rangle$ (all $A_{t,s} \in \mathbf{N}$) commutes with all matrices $\langle B_{t,s}\rangle \in \mathfrak{Z}$. For this it is necessary that it commute with all finite matrices $\langle B_{t,s}\rangle$ (all $B_{t,s} \in \mathbf{N}$, cf. B_p)). This is immediately seen to imply that $A_{t,s} = \delta_{t,s}A$ $(t, s \leqq p)$ for an A belonging to the center of **N**. This condition is also clearly sufficient. Hence the A^0 of the center of **M** are precisely the $\alpha 1$ corresponding to the matrices $\langle \alpha\delta_{t,s}1\rangle$ if and only if the A of the center of **N** are precisely the $\alpha 1$, i.e. **M** is a factor if and only if **N** is one.

This completes the proof.

Ad (ii). By Theorem V, **N** is isomorphic to an $\mathbf{M}_{(\mathfrak{M})}$. Hence our assertion follows from [5], p. 189, Lemma 11.4.3.

Ad (iii). The $W_{\mu,\mu}$ $(\mu \leqq p)$ are pairwise orthogonal projections $\in \mathbf{M}$ all with the same relative dimension in **M** owing to (i), (ii) in Lemma 2.6.1. Hence

$$(2.7.\alpha) \qquad D_{\mathbf{M}}(1) = \sum_{\mu=1}^{p} D_{\mathbf{M}}(W_{\mu,\mu}) = pD_{\mathbf{M}}(W_{1,1}).^{16}$$

[16] Here as in some subsequent instances (proof of Lemma 5.3.4; III in §5.5) it is convenient to use the convention $\infty \cdot 0 = 0$.

Now $W_{1,1}$ is not zero, since if it were we should have $W_{\mu,\mu} = 0$ and $\sum_{\mu=1}^{p} W_{\mu,\mu} = 0$ for $p = 1, 2, \cdots, \infty$. Hence $D_{\mathfrak{M}}(W_{1,1}) \neq 0$. Now $\mathbf{M}$ is in a finite case if and only if $D_{\mathbf{M}}(1)$ is finite. By the above, this is equivalent to the statement that both p and $D_{\mathbf{M}}(W_{1,1})$ are finite. And the finiteness of $D_{\mathbf{M}}(W_{1,1}) = D_{\mathbf{M}}(\mathfrak{M})$, $(P_{\mathfrak{M}} = W_{1,1})$ is, according to [5], p. 189, Lemma 11.4.3, equivalent to $\mathbf{M}_{(\mathfrak{M})}$ being in a finite case, i.e. by Theorem V, to $\mathbf{N}$ being in a finite case.

This completes the proof.

§2.8 Let $\mathbf{M}$ be a factor in a Hilbert space $\mathfrak{H}$. As usual we denote by $D_{\mathbf{M}}(E)$ and by $D_{\mathbf{M}}(\mathfrak{M})$, $(E = P_{\mathfrak{M}} \in \mathbf{M})$ a fixed relative dimension function of $\mathbf{M}$. We consider again the rings $\mathbf{M}^{E}$ and $\mathbf{M}_{(\mathfrak{M})}$ as described in the beginning of §2.5.

LEMMA 2.8.1. *If $\mathfrak{M}_1$, $\mathfrak{M}_2$ are two closed linear sets η $\mathbf{M}$ with $D_{\mathbf{M}}(\mathfrak{M}_1) = D_{\mathbf{M}}(\mathfrak{M}_2)$ then $\mathbf{M}_{(\mathfrak{M}_1)}$ and $\mathbf{M}_{(\mathfrak{M}_2)}$ (in $\mathfrak{M}_1$ and $\mathfrak{M}_2$, respectively) are spatially isomorphic.*

PROOF. Since $D_{\mathbf{M}}(\mathfrak{M}_1) = D_{\mathbf{M}}(\mathfrak{M}_2)$ there exists a partially isometric $U \in \mathbf{M}$ with the initial set $\mathfrak{M}_1$ and final set $\mathfrak{M}_2$. This is a one-to-one isomorphic mapping of $\mathfrak{M}_1$ on $\mathfrak{M}_2$ and it obviously carries $\mathbf{M}_{(\mathfrak{M}_1)}$ into $\mathbf{M}_{(\mathfrak{M}_2)}$. Hence it establishes the desired spatial isomorphism.

For the remainder of this chapter, we assume that the factor $\mathbf{M}$ is in a finite case.

Consider a real number α and a projection $E \in \mathbf{M}$ or equivalently a closed linear set $\mathfrak{M}$ η $\mathbf{M}$ with $E = P_{\mathfrak{M}}$. We assume $E \neq 0$, i.e. $\mathfrak{M} \neq \{0\}$ and

$$(2.8.\alpha) \qquad \alpha = \frac{D_{\mathbf{M}}(E)}{D_{\mathbf{M}}(1)}.$$

Then the above Lemma 2.8.1 shows that $\overline{\mathbf{M}_{(\mathfrak{M})}}$ depends only on $\mathbf{M}$ and α but not on the E, $\mathfrak{M}$ themselves. (A spatial isomorphism implies an algebraic isomorphism.)

Consider now two algebraically isomorphic $\mathbf{M}_1$ and $\mathbf{M}_2$. Choose $E_1 \in \mathbf{M}$ with $\alpha = D_{\mathbf{M}_1}(E)/D_{\mathbf{M}_1}(1)$. Then the isomorphism carries E_1 into an $E_2 \in \mathbf{M}$ with $\alpha = D_{\mathbf{M}_2}(E_2)/D_{\mathbf{M}_2}(1)$. Let $\mathfrak{M}_1$, $\mathfrak{M}_2$ be η $\mathbf{M}_1$, η $\mathbf{M}_2$ respectively, and such that $P_{\mathfrak{M}_1} = E_1$; $P_{\mathfrak{M}_2} = E_2$. The isomorphism carries $\mathbf{M}_1^{E_1}$ into $\mathbf{M}_2^{E_2}$ in particular E_1 into E_2 and it leaves the operations αA, A^*, $A + B$, AB invariant. Hence it generates, by Lemma 2.5.1, an algebraic isomorphism of $(\mathbf{M}_1)_{(\mathfrak{M}_1)}$ and $(\mathbf{M}_2)_{(\mathfrak{M}_2)}$. Thus $\overline{\mathbf{M}_{(\mathfrak{M})}}$ (assuming its existence) is uniquely determined by $\overline{\mathbf{M}}$ and α. We denote this $\overline{\mathbf{M}_{(\mathfrak{M})}}$ by $\overline{\mathbf{M}}^{\alpha}$.

This notation could conflict with that of Theorem IV when $p = \alpha = 1$ But in that case clearly $\overline{\mathbf{M}}^1$ exists and is equal to $\overline{\mathbf{M}}$ under both definitions.

The α, for which $\overline{\mathbf{M}}^{\alpha}$ can be formed, comprise precisely the range of $D_{\mathbf{M}}(E)/D_{\mathbf{M}}(1)$ for all $E \in \mathbf{M}$, $E \neq 0$. This is the set $(1/n, 2/n, \cdots, 1)$ if $\overline{\mathbf{M}}$ is in a case I_n and the set $0 < \alpha \leqq 1$ if $\overline{\mathbf{M}}$ is in a case (II_1).

Summing up,

THEOREM VI. *$\overline{\mathbf{M}}^{\alpha}$ exists if and only if α belongs to the following set: $(1/n, 2/n, \cdots, 1)$ if $\mathbf{M}$ is in a case (I_n), $(n = 1, 2, \cdots)$, $(0 < \alpha \leqq 1)$ if $\mathbf{M}$ is in the case (II_1).*

There is no conflict between this notation and that of Theorem IV.

Since the relative dimension $D_{\mathbf{M}}(E)$ can be characterized in a way which is

invariant under conjugate isomorphisms (cf. the remark at the end of §2.3), our present definition and that of §2.3 yield together

(2.8.β) $$(\overline{\mathbf{M}}^c)^\alpha = (\overline{\mathbf{M}}^\alpha)^c \qquad (\alpha \text{ in the range of Theorem VI}).$$

Consider next two projections $E, F \in \mathbf{M}$, $E, F \neq 0$ and $E \leqq F$ or equivalently the closed linear sets $\mathfrak{M}, \mathfrak{N} \ \eta\ \mathbf{M}$ with $P_{\mathfrak{M}} = E, P_{\mathfrak{N}} = F$. Hence $\mathfrak{M} \neq (0), \mathfrak{N} \neq (0)$ and $\mathfrak{M} \subseteq \mathfrak{N}$. To avoid misunderstanding, let 1 be the unit operator in $\mathfrak{H}$ and $1' = F_{(\mathfrak{N})}$ be the unit in $\mathfrak{N}$. Put

$$\alpha = \frac{D_{\mathbf{M}}(F)}{D_{\mathbf{M}}(1)}, \qquad \beta = \frac{D_{\mathbf{M}_{(\mathfrak{N})}}(E_{(\mathfrak{N})})}{D_{\mathbf{M}_{(\mathfrak{N})}}(1')} = \frac{D_{\mathbf{M}}(E)}{D_{\mathbf{M}}(F)}.$$

Clearly $\alpha\beta = D_{\mathbf{M}}(E)/D_{\mathbf{M}}(1)$ and $\mathbf{M}_{(\mathfrak{M})} = (\mathbf{M}_{(\mathfrak{N})})_{(\mathfrak{M})}$. Our definitions now yield immediately

(2.8.γ) $$(\overline{\mathbf{M}}^\alpha)^\beta = \overline{\mathbf{M}}^{\alpha\beta} \qquad (\alpha, \beta \text{ in the range of Theorem VI}).$$

§2.9 We continue the discussion of the preceding section

LEMMA 2.9.1. *If p is finite and* $\mathbf{M}$ *and* $\mathbf{N}$ *are factors in a finite case,* $\overline{\mathbf{N}} = \overline{\mathbf{M}}^{1/p}$ *is equivalent to the statement* $\overline{\mathbf{M}} = \overline{\mathbf{N}}^p$. *(The statement "*$\overline{\mathbf{M}}^{1/p} = \overline{\mathbf{N}}$*" includes "*$\overline{\mathbf{M}}^{1/p}$ *exists.")*

PROOF. Assume $\overline{\mathbf{M}} = \overline{\mathbf{N}}^p$. It follows from Theorem V that there is an E with $D_{\mathbf{M}}(E) = (1/p)D_{\mathbf{M}}(1)$ such that $\overline{\mathbf{N}} = \overline{\mathbf{M}}_{(\mathfrak{M})}$ where $\mathfrak{M}$ is such that $E = P_{\mathfrak{M}}$. Hence, by the above definition, $\overline{\mathbf{M}}^{1/p} = \overline{\mathbf{N}}$.

Conversely, suppose $\overline{\mathbf{M}}^{1/p} = \overline{\mathbf{N}}$. It follows that $1/p$ is in the set of Theorem VI and hence, if $\mathbf{M}$ is in a case (I_n) p divides n. Thus Lemmas 2.6.2 and 2.6.3 imply that there is an $\overline{\mathbf{N}}_0$ such that $\overline{\mathbf{N}}_0^p = \overline{\mathbf{M}}$. Hence the above result (with $\overline{\mathbf{N}}_0$ in place of $\overline{\mathbf{N}}$) gives $\overline{\mathbf{N}}_0 = \overline{\mathbf{M}}^{1/p}$ i.e. $\overline{\mathbf{N}}_0 = \overline{\mathbf{N}}$. Consequently $\overline{\mathbf{N}}^p = \overline{\mathbf{N}}_0^p = \overline{\mathbf{M}}$.

This completes the proof.

LEMMA 2.9.2. If $\overline{\mathbf{M}}^\alpha$ *exists (in the sense of Theorem* VI) *and if* $p\alpha = q\beta$, $(p, q = 1, 2, \cdots)$ *then* $(\overline{\mathbf{M}}^\alpha)^p = (\overline{\mathbf{M}}^q)^\beta$. *(This is to include the statement that* $(\overline{\mathbf{M}}^q)^\beta$ *exists.)*

PROOF. $p\alpha = q\beta$. Hence $\alpha/q = \beta/p$, $\overline{\mathbf{M}}^\alpha$ exists. From this we shall draw inferences so that the existence of each expression that appears will be established when we write it down. In this way (2.8.γ) and Lemma 2.9.1 give

$$\overline{\mathbf{M}}^\alpha = ((\overline{\mathbf{M}}^q)^{\frac{1}{q}})^\alpha = (\overline{\mathbf{M}}^q)^{\frac{\alpha}{q}} = (\overline{\mathbf{M}}^q)^{\frac{\beta}{p}}$$

Hence by Lemma 2.9.1,

$$(\overline{\mathbf{M}}^\alpha)^p = ((\overline{\mathbf{M}}^q)^{\frac{\beta}{p}})^p = (((\overline{\mathbf{M}}^q)^\beta)^{\frac{1}{p}})^p = (\overline{\mathbf{M}}^q)^\beta$$

LEMMA 2.9.3. *For a given* $\overline{\mathbf{M}}$ *consider all* α *of the set in Theorem* VI, *and all* $p = 1, 2, \cdots$. *Then we have*

(i) *The range of* $\theta = p\alpha$ *is the set* $(1/n, 2/n, \cdots)$ *if* $\mathbf{M}$ *is in a case* (I_n) $(n = 1, 2, \cdots)$ *and the set* $0 < \theta < \infty$ *if* $\mathbf{M}$ *is in the case* (II_1).

(ii) $(\overline{\mathbf{M}}^{\alpha})^{p}$ *depends only on* $\theta = p\alpha$ *and not on the individual values of* α *and* p.

PROOF. Ad (i). Immediate by Theorem V.

Ad (ii). Lemma 2.9.2 yields, when $p\alpha = q\beta$, $(\overline{\mathbf{M}}^{\alpha})^{p} = (\overline{\mathbf{M}}^{q})^{\beta}$ and $(\overline{\mathbf{M}}^{\beta})^{q} = (\overline{\mathbf{M}}^{q})^{\beta}$. Hence $(\overline{\mathbf{M}}^{\alpha})^{p} = (\overline{\mathbf{M}}^{\beta})^{q}$.

We denote the above $(\overline{\mathbf{M}}^{\alpha})^{p}$, $\theta = p\alpha$ by $\overline{\mathbf{M}}^{\theta}$.

This notation does not conflict with those of Theorems IV and VI. This can be seen by putting $\alpha = 1$ or $p = 1$, respectively.

Summing up,—

THEOREM VII. $\overline{\mathbf{M}}^{\theta}$ *exists if and only if* θ *belongs to the following set*: $(1/n, 2/n, \cdots)$ *if* $\mathbf{M}$ *is in a case* (I_n), $(n = 1, 2, \cdots)$, $(0 < \theta < \infty)$ *if* $\mathbf{M}$ *is in the case* (II_1).

There is no conflict between this notation and that of Theorems IV and VI.

Combining $(2.4.\alpha)$ and $(2.8.\beta)$ gives

$$(2.9.\alpha) \qquad (\overline{\mathbf{M}}^{c})^{\theta} = (\overline{\mathbf{M}}^{\theta})^{c} \qquad (\theta \text{ in the set of Theorem VII}).$$

Also, by the use of $(2.4.\beta)$, $(2.8.\gamma)$ and Lemma 2.9.2, we obtain

$$(((\overline{\mathbf{M}}^{\alpha})^{p})^{\beta})^{q} = (((\overline{\mathbf{M}}^{\alpha})^{\beta})^{p})^{q} = (\overline{\mathbf{M}}^{\alpha\beta})\mathbf{M}^{pq}.$$

Hence if we put $\theta = p\alpha$, $\xi = q\beta$, then

$$(2.9.\beta) \qquad (\overline{\mathbf{M}}^{\theta})^{\xi} = (\overline{\mathbf{M}}^{\xi})^{\theta} = \overline{\mathbf{M}}^{\xi\theta} \qquad (\theta \text{ and } \xi \text{ in the set of Theorem VII}).$$

(Concerning the exponent ∞, consider the remark at the end of §2.5.)

§2.10. Denote the set of all θ for which $\overline{\mathbf{M}}^{\theta}$ exists and equals $\overline{\mathbf{M}}$ by

$$(2.10.\alpha) \qquad \mathfrak{G} = \mathfrak{G}(\overline{\mathbf{M}}).$$

Obviously $1 \in \mathfrak{G}$ and $\theta, \xi \in \mathfrak{G}$ imply $\theta\xi \in \mathfrak{G}$. Since $(\overline{\mathbf{M}}^{\theta})^{1/\theta}$ exists and equals $\overline{\mathbf{M}}$, $\overline{\mathbf{M}}^{\theta} = \overline{\mathbf{M}}$ implies $\overline{\mathbf{M}}^{1/\theta} = \overline{\mathbf{M}}$ i.e. $\theta \in \mathfrak{G}$ implies $1/\theta \in \mathfrak{G}$. (All this is due to $(2.9.\beta)$ in §2.9.)

THEOREM VIII. *The set* $\mathfrak{G}$ *of* $(2.10.\alpha)$ *above is a subgroup of* P *the multiplication group of the real numbers* $\theta, 0 < \theta < \infty$. *We call* $\mathfrak{G}$ *the fundamental group of* $\overline{\mathbf{M}}$.

Some properties of $\mathfrak{G}$.

LEMMA 2.10.1. $\overline{\mathbf{M}}^{\theta} = \overline{\mathbf{M}}^{\xi}$ (*both sides are assumed to exist*) *is equivalent to* $\theta/\xi \in \mathfrak{G}$.

PROOF. Sufficiency: $\theta/\xi \in \mathfrak{G}$ implies $\overline{\mathbf{M}}^{\theta/\xi} = \overline{\mathbf{M}}$. Hence $\overline{\mathbf{M}}^{\theta} = (\overline{\mathbf{M}}^{\theta/\xi})^{\xi} = \overline{\mathbf{M}}^{\xi}$.

Necessity: Assume $\overline{\mathbf{M}}^{\theta} = \overline{\mathbf{M}}^{\xi}$. $(\overline{\mathbf{M}}^{\xi})^{1/\xi}$ exists and equals $\overline{\mathbf{M}}$. Hence $(\overline{\mathbf{M}}^{\theta})^{1/\xi} = \overline{\mathbf{M}}^{\theta/\xi}$ exists and equals $\overline{\mathbf{M}}$. Thus $\theta/\xi \in \mathfrak{G}$.

LEMMA 2.10.2. $\overline{\mathbf{M}}$, $\overline{\mathbf{M}}^{c}$ *and all* $\overline{\mathbf{M}}^{\theta}$ *have the same fundamental group.*

PROOF. Ad $\overline{\mathbf{M}}^{c}$. Immediate by $(2.9.\alpha)$.

Ad. $\overline{\mathbf{M}}^{\theta}$. Immediate by Lemma 2.10.1.

Before we conclude this chapter, we determine the behavior of the discrete finite cases, i.e. the (I_n), $n = 1, 2, \cdots$.

Consider the ring $\mathbf{\Theta}$ of all operators $\alpha \cdot 1$. This is obviously the prototype of all factors of the case (I_1), its type $\overline{\mathbf{\Theta}}$ is that of the ring of all complex numbers.

Now we have

LEMMA 2.10.3. *If* $\mathbf{M}$ *is in case* (I_n), $(n = 1, 2, \cdots)$ *then* $\overline{\mathbf{M}}^\alpha$ *(α finite), exists if and only if* $n\alpha = 1, 2, \cdots$, *and then* $\overline{\mathbf{M}}^\alpha = \overline{\Theta}^{n\alpha}$.

PROOF. The range asserted for α is the same as that of Theorem VII. Put $n\alpha = k$. Then

$$\overline{\mathbf{M}}^\alpha = \overline{\mathbf{M}}^{\frac{k}{n}} = \left(\overline{\mathbf{M}}^{\frac{1}{n}}\right)^k.$$

Now consider $\overline{\mathbf{M}}^{1/n}$. Apply the definition of §2.8, preceding Theorem VII. For the E in question, $D_{\mathbf{M}}(E)/D_{\mathbf{M}}(1)$ assumes its minimal value, $1/n$. Hence, this E is minimal. (Cf. [5], pp. 143–144, Def. 5.1.2.) Consequently $\mathbf{M}^E$ is the set of all αE (cf. [5], p. 144, the last part of the proof of Lemma 5.1.3—or it may be proved directly). Hence $\mathbf{M}_{(\mathfrak{M})}$ is the set of all $\alpha \cdot 1$ in $\mathfrak{M}$ i.e.

$$\overline{\mathbf{M}}^{\frac{1}{n}} = \overline{\mathbf{M}}_{(\mathfrak{M})} = \overline{\Theta}.$$

Consequently

$$\overline{\mathbf{M}}^\alpha = \overline{\Theta}^k, \qquad k = n\alpha$$

as desired.

COROLLARY. *If* $\mathbf{M}$ *is in a case* $(I_n)(n = 1, 2, \cdots)$ *then* $\overline{\mathbf{M}} = \Theta^n$.

PROOF. Put $\alpha = 1$ in the above lemma.

LEMMA 2.10.4. *If* $\mathbf{M}$ *is in a case* $(I_n)(n = 1, 2, \cdots)$ *then* $\overline{\mathbf{M}}^c = \overline{\mathbf{M}}$ *and* $\mathfrak{G} = \mathfrak{G}(\mathbf{M}) = (1)$.

PROOF.[17] Ad. $\overline{\mathbf{M}}^c = \overline{\mathbf{M}}$. Clearly $\overline{\Theta}^c = \Theta$. Hence the corollary to Lemma 2.10.3, and (2.4.α) or (2.9.α) imply $\overline{\mathbf{M}}^c = \overline{\mathbf{M}}$.

Ad. $\mathfrak{G} = (1)$. By Lemma 2.10.3 and its corollary, $\overline{\mathbf{M}}^\alpha$ is in case (I_k), $k = n\alpha$. Hence $\overline{\mathbf{M}}^\alpha = \overline{\mathbf{M}}$ implies $n\alpha = n$ or $\alpha = 1$. So $\mathfrak{G} = (1)$.

The validity or invalidity of the equation $\overline{\mathbf{M}}^c = \overline{\mathbf{M}}$ and the fundamental group $\mathfrak{G} = \mathfrak{G}(\overline{\mathbf{M}})$ are algebraical isomorphism invariants of $\mathbf{M}$. Lemma 2.10.4 showed us how they behave when $\mathbf{M}$ is in a discrete finite case, i.e. in a case (I_n), $n = 1, 2, \cdots$. The really interesting factors are the remaining finite ones, those in the continuous finite case (II_1). We shall see that they are not all isomorphic to each other, but we shall achieve this differentiation with the help of other invariants. (Cf. the end of §5.6 and the beginning of §6.1.)

For all $\overline{\mathbf{M}}$ in case (II_1) for which we have succeeded in settling this question, $\overline{\mathbf{M}}^c = \overline{\mathbf{M}}$ and $\overline{\mathbf{M}}^\theta = \overline{\mathbf{M}}$ (θ in the set of Theorem VII), i.e. $\mathfrak{G} = P$, was found. (Cf. Theorem XV and the end of §5.6. Observe Lemma 2.10.4 by comparison.) There seems to be, however, no reason to believe the general validity of these relations. The general behavior of the above invariants remains therefore an open question.

CHAPTER III. CHARACTERIZATION OF ALL SPATIAL TYPES IN TERMS OF ALGEBRAICAL TYPES

§3.1 **We have classified all operator rings $\mathbf{M}$ according to their types $\overline{\mathbf{M}}$ i.e. with respect to algebraical isomorphism. Now we shall introduce a broader**

[17] **Both assertions could be proved directly by matrix considerations.**

classification, to be called the *genus*, each genus consisting of one or more types. It will be necessary, however, to restrict this new classification to factors $\mathbf{M}$.

We need some auxiliary lemmas which lead up to the desired definition. Throughout this section $\mathbf{M}$ will be a factor in a Hilbert space $\mathfrak{H}$. As before, we denote by $D_{\mathbf{M}}(E)$ and $D_{\mathbf{M}}(\mathfrak{M})$ a fixed relative dimension function of $\mathbf{M}$ applicable to the $E \in \mathbf{M}$ or the $\mathfrak{M} \,\eta\, \mathbf{M}$.

LEMMA 3.1.1. *If $\mathfrak{M}$ is a closed linear set, $\mathfrak{M} \,\eta\, \mathbf{M}$ and such that $\mathfrak{M} \neq (0)$ and $D_{\mathbf{M}}(\mathfrak{H} \ominus \mathfrak{M}) = \infty$ then*

$$\bar{\mathbf{M}} = (\overline{\mathbf{M}_{(\mathfrak{M})}})^{\infty}.$$

PROOF. Under our hypothesis, $\mathfrak{M}$ "divides" $\mathfrak{H} \ominus \mathfrak{M}$ an infinite number of times, i.e. there are an infinite number of pairwise orthogonal closed linear sets $\mathfrak{M}_2'$, $\mathfrak{M}_3'$, $\cdots$ such that $D_{\mathbf{M}}(\mathfrak{M}_i') = D_{\mathbf{M}}(\mathfrak{M})$ and $\mathfrak{M}_i' \subset \mathfrak{H} \ominus \mathfrak{M}$ for $i = 2, 3, \cdots$. If $\mathfrak{M}$ is relatively finite-dimensional, this is obvious. If $\mathfrak{M}$ is infinite-dimensional, we apply a variant of [5], Lemma 7.2.3, p. 157, to $\mathfrak{H} \ominus \mathfrak{M}$ which shows that $\mathfrak{H} \ominus \mathfrak{M}$ is determined by a denumerably infinite number of pairwise orthogonal closed linear sets $\mathfrak{M}_2'$, $\mathfrak{M}_3'$, $\cdots$, each of which is relatively infinite-dimensional. (To prove the variant, it is only necessary, in the proof of [5], Lemma 7.2.3, to divide the $\mathfrak{P}_1$, $\mathfrak{P}_2$, $\cdots$, into an infinite number of sets, each having an infinite number of $\mathfrak{P}_i$'s, rather than into two sets.)

If we now apply [5], Lemma 7.1.2, p. 155, to $\mathfrak{M}$ and $\mathfrak{H} \ominus \mathfrak{M}$ we see that $\mathfrak{H} \ominus \mathfrak{M}$ is determined by a denurably infinite number of sets $\mathfrak{M}_2$, $\mathfrak{M}_3$, $\cdots$, which are pairwise orthogonal and for which $D_{\mathbf{M}}(\mathfrak{M}_i) = D_{\mathbf{M}}(\mathfrak{M})$ for $i = 2, 3, \cdots$. Let $\mathfrak{M} = \mathfrak{M}_1$.

The second paragraph of the proof of Lemma 2.6.2 can now be used to show the existence in $\mathbf{M}$ of a set of ∞-matrix units $W_{\mu,\nu}$ with $W_{\mu,\mu} = P_{\mathfrak{M}_\mu}$, $W_{1,1} = P_{\mathfrak{M}_1} = P_{\mathfrak{M}}$. For the hypothesis $p < \infty$ is not used here. Furthermore, $E_0 = \sum_{\mu=1}^{\infty} W_{\mu,\mu} = W_{1,1} + \sum_{\mu=2}^{\infty} W_{\mu,\mu} = P_{\mathfrak{M}_1} + \sum_{\mu=2}^{\infty} P_{\mathfrak{M}_\mu} = P_{\mathfrak{M}} + P_{\mathfrak{H}\ominus\mathfrak{M}} = 1$. Thus Theorem V in §2.6 above yields the desired result.

LEMMA 3.1.2. *If $\mathfrak{M}$ is a closed linear set, $\mathfrak{M} \,\eta\, \mathbf{M}$ and $\neq (0)$ and if $\mathbf{M}$ is in an infinite case, then*

$$\bar{\mathbf{M}} = (\overline{\mathbf{M}_{(\mathfrak{M})}})^{\infty}.$$

PROOF. If $D_{\mathbf{M}}(\mathfrak{H} \ominus \mathfrak{M}) = \infty$ the result is implied by Lemma 3.1.1. Hence we may assume that $D_{\mathbf{M}}(\mathfrak{H} \ominus \mathfrak{M})$ is finite. Since $\mathbf{M}$ is in an infinite case, $D_{\mathbf{M}}(\mathfrak{H})$ is infinite and hence $D_{\mathbf{M}}(\mathfrak{M}) = D_{\mathbf{M}}(\mathfrak{H}) - D_{\mathbf{M}}(\mathfrak{H} \ominus \mathfrak{M})$ is infinite also.

Since $D_{\mathbf{M}}(\mathfrak{H})$ is infinite, there exists by [5], Lemma 2.7.3, p. 157, a closed linear set $\mathfrak{N}$ such that $D_{\mathbf{M}}(\mathfrak{N})$ and $D_{\mathbf{M}}(\mathfrak{H} \ominus \mathfrak{N})$ are also infinite. Hence $\overline{\mathbf{M}_{(\mathfrak{M})}} = \overline{\mathbf{M}_{(\mathfrak{N})}}$ by lemma 2.8.1 (we need only the algebraical, not the spatial, isomorphism) and $\bar{\mathbf{M}} = (\overline{\mathbf{M}_{(\mathfrak{N})}})^{\infty}$ by Lemma 3.1.1. These give together $\bar{\mathbf{M}} = (\overline{\mathbf{M}_{(\mathfrak{M})}})^{\infty}$.

LEMMA 3.1.3. *A factor $\mathbf{M}$ is in an infinite case if and only if*

$$\bar{\mathbf{M}} = \bar{\mathbf{M}}^{\infty}.$$

PROOF. Sufficiency: Immediate by (iii) in Lemma 2.7.1.
Necessity: Put $\mathfrak{M} = \mathfrak{H}$ in Lemma 3.1.2.

DEFINITION 3.1.1. *If, for two factors,* **M**, **N**

$$\bar{\mathbf{N}} = \overline{\mathbf{M}_{(\mathfrak{M})}}$$

for a suitable closed set $\mathfrak{M} \neq (0)$ *and* $\mathfrak{M}\ \eta\ \mathbf{M}$ *in the Hilbert space* $\mathfrak{H}$ *of* **M** *then we say that* $\bar{\mathbf{M}}$ *is a multiple of* $\bar{\mathbf{N}}$, *or that* $\bar{\mathbf{N}}$ *is a divisor of* $\bar{\mathbf{M}}$.

It is clear that this relationship concerns $\bar{\mathbf{M}}$ and $\bar{\mathbf{N}}$. Cf. the argument immediately preceding Theorem VI in §2.8. It is also obviously transitive.

LEMMA 3.1.4. $\bar{\mathbf{N}}$ *is a divisor of* $\bar{\mathbf{M}}$ *if and only if* $\alpha)\bar{\mathbf{M}} = \bar{\mathbf{N}}$ *when* $\bar{\mathbf{N}}$ *is in an infinite case* $\beta)$ $\bar{\mathbf{M}} = \bar{\mathbf{N}}^\theta$ *with a* $\theta \geqq 1$ *in the range of Theorem* VII *or of Theorem* IV′ (i.e., $\theta = \infty$) *when* $\bar{\mathbf{N}}$ *is in a finite case.*

PROOF. Ad α). Sufficiency: Obvious.

Necessity: Since $\bar{\mathbf{N}} = \overline{\mathbf{M}_{(\mathfrak{M})}}$ is in an infinite case, $D_{\mathbf{M}}(\mathfrak{M})$ is infinite by [5], p. 189, Lemma 11.4.3. Hence **M** is in an infinite case and thus Lemma 3.1.2 gives $\bar{\mathbf{M}} = (\overline{\mathbf{M}_{(\mathfrak{M})}})^\infty = \bar{\mathbf{N}}^\infty$. On the other hand, $\bar{\mathbf{N}} = \bar{\mathbf{N}}^\infty$ by Lemma 3.1.3. Thus $\bar{\mathbf{M}} = \bar{\mathbf{N}}$.

Ad β). Sufficiency: θ finite: Then $\bar{\mathbf{N}} = \bar{\mathbf{M}}^\alpha$ with $\alpha = 1/\theta < 1$. Hence the assertion follows by our definition of $\bar{\mathbf{M}}^\alpha$ for $\alpha \leqq 1$ in §2.8. θ infinite: Immediate by Theorem V.

Necessity: $\bar{\mathbf{M}}$ in a finite case: Clearly $\bar{\mathbf{N}} = \overline{\mathbf{M}_{(\mathfrak{M})}} = \bar{\mathbf{M}}^\alpha$ with $\alpha = D_{\mathbf{M}}(E)/D_{\mathbf{M}}(1) \leqq 1$, $(E = P_{\mathfrak{M}})$. Hence $\bar{\mathbf{M}} = \bar{\mathbf{N}}^\theta$ with $\theta = 1/\alpha \geqq 1$. $\bar{\mathbf{M}}$ in an infinite case: $\bar{\mathbf{M}} = (\overline{\mathbf{M}_{(\mathfrak{M})}})^\infty = \bar{\mathbf{N}}^\infty$ by Lemma 3.1.2.

LEMMA 3.1.5. *The four following statements are equivalent to each other:*

α) $\bar{\mathbf{M}}$ *is either a multiple or divisor of* $\bar{\mathbf{N}}$,

β) $\bar{\mathbf{M}}$ *and* $\bar{\mathbf{N}}$ *possess a common multiple,*

γ) $\bar{\mathbf{M}}$ *and* $\bar{\mathbf{N}}$ *possess a common divisor,*

δ) $\bar{\mathbf{M}}^\infty = \bar{\mathbf{N}}^\infty$.

PROOF. The implications

$$\alpha) \to \beta), \qquad \alpha) \to \gamma)$$

are obvious. Furthermore

$$\delta) \to \beta)$$

follows from Theorem V, since this theorem shows that $\bar{\mathbf{M}}^\infty = \bar{\mathbf{N}}^\infty$ is a multiple of $\bar{\mathbf{M}}$ and of $\bar{\mathbf{N}}$.

We prove next

$$\beta) \to \alpha)$$

Let $\bar{\mathbf{L}}$ be the common multiple of both $\bar{\mathbf{M}}$ and $\bar{\mathbf{N}}$. Let closed linear sets $\mathfrak{M}$ and $\mathfrak{N}$ be chosen in the $\mathfrak{H}$ associated with **L**, so that $\bar{\mathbf{M}} = \overline{\mathbf{L}_{(\mathfrak{M})}}$ and $\bar{\mathbf{N}} = \overline{\mathbf{L}_{(\mathfrak{N})}}$. Since **L** is a factor either $\mathfrak{M}$ has the same relative dimension (in **L**) as an $\mathfrak{N}' \subseteq \mathfrak{N}$ or $\mathfrak{N}$ has the same relative dimension as an $\mathfrak{M}' \subset \mathfrak{M}$. Cf. [5], p. 153, Lemma 6.2.3, or p. 154, Theorem VI. By symmetry we may assume the former. By Lemma 2.8.1 we may now replace $\mathfrak{M}$ by $\mathfrak{N}'$, *i.e.* we can assume that $\mathfrak{M} \subseteq \mathfrak{N}$.

Thus $(\overline{\mathbf{L}_{(\mathfrak{N})(\mathfrak{M})}}) = \overline{\mathbf{L}_{(\mathfrak{M})}}$ so that $\overline{\mathbf{L}_{(\mathfrak{N})}}$ is a multiple of $\overline{\mathbf{L}_{(\mathfrak{M})}}$. Hence $\overline{\mathbf{N}}$ is a multiple of $\overline{\mathbf{M}}$.

Finally

$$\gamma) \rightarrow \delta)$$

Let $\overline{\mathbf{K}}$ be a common divisor of $\overline{\mathbf{M}}$ and $\overline{\mathbf{N}}$. Since $\overline{\mathbf{M}}$ is a divisor of $\overline{\mathbf{M}}^\infty$ by Theorem V, so $\overline{\mathbf{K}}$ is a divisor of $\overline{\mathbf{M}}^\infty$ too. Now $\overline{\mathbf{M}}^\infty$ is in an infinite case by (iii) in Lemma 2.7.1. Hence $\overline{\mathbf{K}}^\infty = \overline{\mathbf{M}}^\infty$ by Lemma 3.1.2. Similarly $\overline{\mathbf{K}}^\infty = \overline{\mathbf{N}}^\infty$. Consequently $\overline{\mathbf{M}}^\infty = \overline{\mathbf{N}}^\infty$.

All these equivalences give together

$$\alpha) \rightarrow \gamma) \rightarrow \delta) \rightarrow \beta) \rightarrow \alpha)$$

thus completing the proof.

DEFINITION 3.1.2. *If the equivalent conditions of Lemma 3.1.5 are satisfied then we say that* $\overline{\mathbf{M}}, \overline{\mathbf{N}}$ *are commensurable.*

δ) of Lemma 3.1.5 implies that the notion of commensurability is reflexive, symmetric and transitive.

LEMMA 3.1.6. $\overline{\mathbf{M}}, \overline{\mathbf{N}}$ *are commensurable if and only if*

α) $\overline{\mathbf{M}} = \overline{\mathbf{N}}$ *when* $\overline{\mathbf{M}}, \overline{\mathbf{N}}$ *are both in infinite cases.*

β) $\overline{\mathbf{M}} = \overline{\mathbf{N}}^\infty$ *when* $\overline{\mathbf{M}}$ *is in an infinite case and* $\overline{\mathbf{N}}$ *in a finite case*

γ) $\overline{\mathbf{M}}^\infty = \overline{\mathbf{N}}$ *when* $\overline{\mathbf{M}}$ *is in a finite case and* $\overline{\mathbf{N}}$ *in an infinite case*

δ) $\overline{\mathbf{M}} = \overline{\mathbf{N}}^\theta$ *with a* θ *in the range of Theorem* VII *when* $\overline{\mathbf{N}}$ *and* $\overline{\mathbf{M}}$ *are both in finite cases.*

PROOF. Ad $\alpha) - \gamma)$. Condition δ of Lemma 3.1.5 for commensurability is $\overline{\mathbf{M}}^\infty = \overline{\mathbf{N}}^\infty$. Lemma 3.1.3 shows that when $\overline{\mathbf{M}}$ is in an infinite case we may substitute $\overline{\mathbf{M}}$ for $\overline{\mathbf{M}}^\infty$ in this condition and correspondingly for $\overline{\mathbf{N}}$. When $\overline{\mathbf{M}}$ and $\overline{\mathbf{N}}$ are both infinite, this yields α) while if only one is infinite we obtain β) or γ)

Ad δ). By Lemma 3.1.4, β), $\overline{\mathbf{M}} = \overline{\mathbf{N}}^\theta$ for $\theta \geqq 1$ in the range of Theorem VII is equivalent to $\overline{\mathbf{N}}$ a divisor of $\overline{\mathbf{M}}$. Since $\overline{\mathbf{M}}$ is finite, θ is finite by Lemma 2.7.1 (iii).

On the other hand, $\overline{\mathbf{M}} = \overline{\mathbf{N}}^\theta$ for a $\theta \leqq 1$ in the range of Theorem VII (for $\overline{\mathbf{N}}$) is equivalent to $\overline{\mathbf{M}}^{1/\theta} = \overline{\mathbf{N}}$ with $1/\theta$ in the range of Theorem VII (for $\mathbf{M}$). (Cf. (2.9.β). That θ is in the correct sets in the discrete cases is readily shown by reference to Lemma 2.10.3 and its corollary.) Lemma 3.1.4 with $\overline{\mathbf{M}}$ and $\overline{\mathbf{N}}$ interchanged shows that this is equivalent to $\overline{\mathbf{N}}$ is a multiple of $\overline{\mathbf{M}}$.

Thus $\overline{\mathbf{M}} = \overline{\mathbf{N}}^\theta$ for a θ in the range of Theorem VII is equivalent to $\overline{\mathbf{N}}$ is a divisor of $\overline{\mathbf{M}}$ or $\overline{\mathbf{N}}$ is a multiple of $\mathbf{M}$, i.e. to the condition of Lemma 3.1.5 for commensurability.

We call the abstraction of type $\overline{\mathbf{M}}$ (for factors) with respect to commensurability its *genus*, i.e. two types $\overline{\mathbf{M}}_1$ and $\overline{\mathbf{M}}_2$ will be said to have the same genus if and only if they are commensurable. We denote the genus of the type $\overline{\mathbf{M}}$ by $\overline{\overline{\mathbf{M}}}$. Notions invariant under commensurability will be said to be notions concerning the genus $\overline{\overline{\mathbf{M}}}$ (and not $\overline{\mathbf{M}}$ or $\mathbf{M}$ itself).

Thus the cases I, II, III are notions concerning the genus. This follows from (ii) in Lemma 2.7.1 (with $p = \infty$) and the characterization δ) in Lemma 3.1.5.

Similarly $\bar{\mathbf{M}}^c$ is an operation concerning the genus. This follows from (2.5. ϵ and the above-mentioned characterization. Therefore we can define

$$\bar{\bar{\mathbf{M}}}^c = \overline{\overline{\mathbf{M}}^c}.$$

Observe finally that $(\overline{\mathbf{M}_{(\mathfrak{M})}})$, $\bar{\mathbf{M}}^p$, $p = 1, 2, \cdots, \infty$ and (when it is defined, i.e. for $\mathbf{M}$ in a finite case) $\bar{\mathbf{M}}^\theta$ (θ in the range of Theorem VII) have the same genus as $\mathbf{M}$ (for $(\overline{\mathbf{M}_{(\mathfrak{M})}})$ this follows from α) in Lemma 3.1.5; for $\bar{\mathbf{M}}^p$ this is implied by Theorem V and Lemma 3.1.5 α); and for $\bar{\mathbf{M}}^\theta$ this follows from Lemma 3.1.6δ).)

It is clear then that being in a finite or an infinite case does not concern the genus (cf. Theorem IX for details).

§3.2 Since the cases (I), (II), (III) are notions concerning the genus, each one of these cases is the sum of one or more genera. And of course each genus is the sum of one or more types. These are the details.

THEOREM IX. α) *Case* (I) *consists of exactly one genus. This genus contains exactly one type in an infinite case*: (I_∞); *and for every positive integer M exactly one type* (I_n), *these being all its types in finite cases.*

β) *Case* (II) *consists of one or more genera. Each such genus contains exactly one type in an infinite case* (II_∞) *and one or more types in finite cases* (II_1).

γ) *Case* (III) *consists of one or more genera. Each such genus contains exactly) one type which is, of course, in an infinite case* (III_∞).

PROOF. Ad α). Each one of the cases (I_1), (I_2), $\cdots$, (I_∞) contains exactly one type by [5], p. 173, Lemma 8.6.1. They can obviously all be obtained from the last one by the operations $\mathbf{M}_{(\mathfrak{M})}$. Hence they are all of the same genus as this last by α) of Lemma 3.1.5 and Def. 3.1.1.

Ad β). Consider a genus $\bar{\bar{\mathbf{M}}}$ of case (II). Then there exists an $\mathbf{M}_{(\mathfrak{M})}$ in a *finite* case, for any $\mathfrak{M}$ with $D_{\mathbf{M}}(\mathfrak{M}) < \infty$ will do. Thus $\overline{\mathbf{M}_{(\mathfrak{M})}}$ is finite and it belongs to the given genus of $\mathbf{M}$ by condition α) of Lemma 3.1.5. $\bar{\mathbf{M}}^\infty$ belongs also to this genus for the same reason, since $\bar{\mathbf{M}}$ is a divisor of $\bar{\mathbf{M}}^\infty$ by Theorem V of §2.6. The infinite case in a particular genus is unique by α) of Lemma 3.1.6.

Ad γ). Consider a genus $\bar{\bar{\mathbf{M}}}$ of case (III_∞). Then $\bar{\mathbf{M}}$ must be in an infinite case, and consequently it is unique by α) of Lemma 3.1.6.

Observe that none of the three cases (I), (II) or (III) is empty. Cf. [8], p. 94, §1, in the introduction.

LEMMA 3.2.1. *Consider a genus $\bar{\bar{\mathbf{M}}}$ which contains types $\bar{\mathbf{M}}$ in finite cases, i.e. one in the cases* (I), (II) (*cf. Theorem* IX *above*). *Then the fundamental group $\mathfrak{G}(\bar{\mathbf{M}})$ is a notion concerning the genus $\bar{\bar{\mathbf{M}}}$, i.e. it is the same for all the types $\bar{\mathbf{M}}$ of the genus which are in a finite case.*

PROOF. Immediate by δ) in Lemma 3.1.6 and Lemma 2.10.2.

We shall therefore denote the fundamental group from now on by $\mathfrak{G}(\bar{\bar{\mathbf{M}}})$. Some simple properties of $\mathfrak{G}(\bar{\bar{\mathbf{M}}})$ follow.

LEMMA 3.2.2. *Consider a genus $\bar{\bar{\mathbf{M}}}$ as above. Form the ring* $\mathbf{M}$ *and two closed linear sets*, $\mathfrak{M}, \mathfrak{N} \; \eta \; \mathbf{M}$ *which are both* $\neq (0)$ *and of finite relative dimension. Then we have*

(i) $$\overline{\mathbf{M}_{(\mathfrak{M})}} = \overline{\mathbf{M}_{(\mathfrak{N})}^{D_{\mathbf{M}}(\mathfrak{M})/D_{\mathbf{M}}(\mathfrak{N})}},$$

(ii) $$\overline{\mathbf{M}_{(\mathfrak{M})}} = \overline{\mathbf{M}_{(\mathfrak{N})}} \quad \textit{if and only if} \quad D_{\mathbf{M}}(\mathfrak{M})/D_{\mathbf{M}}(\mathfrak{N}) \in \mathfrak{G}(\bar{\bar{\mathbf{M}}}).$$

PROOF. Ad (i). Suppose first that $D'_{\mathbf{M}}(\mathfrak{M}) \leqq D_{\mathbf{M}}(\mathfrak{N})$.

Here $D_{\mathbf{M}}(\mathfrak{M}) = D_{\mathbf{M}}(\mathfrak{M}')$ for an $\mathfrak{M}' \subset \mathfrak{N}$ and since by Lemma 2.8.1, $\overline{\mathbf{M}_{(\mathfrak{M}')}} = \overline{\mathbf{M}_{(\mathfrak{M})}}$ we may suppose $\mathfrak{M} \subset \mathfrak{N}$. If $\mathfrak{P}$ is a set $\eta\, \mathbf{M}$ and $\subset \mathfrak{N}$, $D_{\mathbf{M}}(\mathfrak{P})$ is a dimension function for $\mathbf{M}_{(\mathfrak{N})}$. Since $\mathfrak{M} \subsetneqq \mathfrak{N}$ by definition, $\overline{\mathbf{M}_{(\mathfrak{M})}} = (\overline{\mathbf{M}_{(\mathfrak{N})}})^{\alpha}$ for $\alpha = D_{\mathbf{M}}(\mathfrak{M})/D_{\mathbf{M}}(\mathfrak{N})$. (Cf. §2.8, $\mathfrak{H}$, $\mathbf{M}$ are replaced by $\mathfrak{N}$, $\mathbf{M}_{(\mathfrak{N})}$.)

On the other hand, if $D_{\mathbf{M}}(\mathfrak{M}) > D_{\mathbf{M}}(\mathfrak{N})$ the above argument yields $(\overline{\mathbf{M}_{(\mathfrak{M})}})^{1/\alpha} = \overline{\mathbf{M}_{(\mathfrak{N})}}$. Equation (2.9.$\beta$) can now be used to show that $\overline{\mathbf{M}_{(\mathfrak{M})}} = (\overline{\mathbf{M}_{(\mathfrak{N})}})^{\alpha}$, suitable use of Lemma 2.10.3 and its corollary being made in the discrete case.

Ad (ii). By the definition of $\mathfrak{G}(\overline{\mathbf{M}_{(\mathfrak{N})}})$, (i) implies $\overline{\mathbf{M}_{(\mathfrak{M})}} = \overline{\mathbf{M}_{(\mathfrak{N})}}$ if and only if $\alpha \in \mathfrak{G}(\overline{\mathbf{M}_{(\mathfrak{N})}})$. But $\mathfrak{G}(\overline{\mathbf{M}_{(\mathfrak{N})}}) = \mathfrak{G}(\bar{\bar{\mathbf{M}}})$ by Lemma 3.2.1.

LEMMA 3.2.3. *Consider a genus $\bar{\bar{\mathbf{M}}}$ as above. Then it contains precisely one type in a finite case if and only if* $\mathfrak{G}(\bar{\bar{\mathbf{M}}}) = P$ (cf. Theorem VIII).

PROOF. Sufficiency: Immediate by δ) in Lemma 3.1.6.

Necessity: Assume $\mathfrak{G}(\bar{\bar{\mathbf{M}}}) \neq P$. Consider an $\bar{\mathbf{M}}$ of this genus in a finite case. If we find a θ not $\in \mathfrak{G}(\bar{\bar{\mathbf{M}}})$ which is in the range of Theorem VII, then $\bar{\mathbf{M}}^{\theta}$ has the same genus, is also in a finite case and yet $\neq \bar{\mathbf{M}}$, thus completing the proof.

Now in case (I), $\mathfrak{G}(\bar{\bar{\mathbf{M}}}) = (1)$ by Lemma 2.10.4. Hence any $\theta = 2, 3, \cdots$ will do. And in case (II), P is the range of Theorem VII, so $\mathfrak{G}(\bar{\bar{\mathbf{M}}}) \neq P$ guarantees the existence of the desired θ.

§3.3 The moment has now come when it is opportune to introduce the notion of a spatial type (cf. footnote [12]). We shall call the abstraction of $\mathbf{M}$ with respect to spatial isomorphism its spatial type. Thus two rings $\mathbf{M}_1$ and $\mathbf{M}_2$ in two Hilbert spaces $\mathfrak{H}_1$ and $\mathfrak{H}_2$, respectively, will be said to have the same spatial type if and only if they are spatially isomorphic. We denote the spatial type of the ring $\mathbf{M}$ by $\ddot{\mathbf{M}}$.

We shall now answer Question II in §2 in the introduction, for the cases (I) and (II)[18], i.e. determine when an algebraic isomorphism determines a spatial one. Thus we shall establish the relationship of $\ddot{\mathbf{M}}$ to $\bar{\mathbf{M}}$. For the cases (I) and (II) we shall determine all spatial types $\ddot{\mathbf{M}}$ in terms of the algebraical types $\bar{\mathbf{M}}$.

The notions $\bar{\mathbf{M}}^{\theta}$, $\mathfrak{G}(\bar{\mathbf{M}})$ as well as the genus $\bar{\bar{\mathbf{M}}}$ will be basic in these considerations.

A spatial isomorphism of $\mathbf{M}_1$ in $\mathfrak{H}_1$ and $\mathbf{M}_2$ in $\mathfrak{H}_2$ (cf. the Definition A in §2 in the Introduction) will also carry $\mathbf{M}'_1$ into $\mathbf{M}'_2$. So the spatial type $\ddot{\mathbf{M}}$ of a ring $\mathbf{M}$ determines not only the algebraical type $\bar{\mathbf{M}}$ but also the algebraical type $\overline{\mathbf{M}'}$. Now $\bar{\mathbf{M}}$ does not determine $\overline{\mathbf{M}'}$ uniquely, not even for a factor $\bar{\mathbf{M}}$[19]. Consequently a discussion of the spatial type $\ddot{\mathbf{M}}$ of a factor $\mathbf{M}$ must begin with an investigation of the relation between $\bar{\mathbf{M}}$ and $\overline{\mathbf{M}'}$. At this point the notion of genus comes in.

---

[18] In the cases (I) of course the answer is known. Nevertheless we include them for the reason given in footnote[1].

[19] The direct factors described in [5], p. 139 or p. 173, Lemma 8.6.1, give elementary examples of this. The exhaustive results in this respect are contained in the last part of our Theorem X.

LEMMA 3.3.1. *For every factor* $\mathbf{M}$ *in the cases* (I) *and* (II)

$$\overline{\overline{\mathbf{M}'}} = \overline{\overline{\mathbf{M}}}^c.$$

PROOF. Due to our assumptions concerning $\mathbf{M}$ there exists a closed linear set $\mathfrak{M} \,\eta\, \mathbf{M}$ which is not $= (0)$ and has a finite $D_{\mathbf{M}}(\mathfrak{M})$. Choose an $f \in \mathfrak{M}$ with $f \neq 0$ and form $\mathfrak{M}_f^{\mathbf{M}'}$, $\mathfrak{M}_f^{\mathbf{M}}$ by [5], p. 143, Def. 5.1.1. These are closed linear sets, clearly $\eta\, \mathbf{M}$ and $\eta\, \mathbf{M}'$ respectively, both $\neq (0)$ and $\mathfrak{M}_f^{\mathbf{M}'} \subseteq \mathfrak{M}$ since $f \in \mathfrak{M} \,\eta\, \mathbf{M}$. Hence $D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}'})$ is finite.

We now normalize $D_{\mathbf{M}}$ so as to make $D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}'}) = 1$ and then $D_{\mathbf{M}'}$ so as to make $C = 1$ for the C of [5], p. 182, Theorem X. Hence

$$D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}'}) = D_{\mathbf{M}'}(\mathfrak{M}_f^{\mathbf{M}}) = 1. \tag{3.3.α}$$

For the sake of brevity we write

$$\mathfrak{M}_1 = \mathfrak{M}_f^{\mathbf{M}'}, \qquad \mathfrak{M}_1' = \mathfrak{M}_f^{\mathbf{M}}. \tag{3.3.β}$$

Let us now use the considerations of [5], pp. 188–190, §11.4. We form in accord with them, $\mathbf{M}_{(\mathfrak{M}_1 \cdot \mathfrak{M}_1')}$, $\mathbf{M}'_{(\mathfrak{M}_1 \cdot \mathfrak{M}_1')}$. These are coupled factors in $\mathfrak{M}_1 \cdot \mathfrak{M}_1' \neq (0)$ just as $\mathbf{M}$, $\mathbf{M}'$ are in $\mathfrak{H}$. Cf. the remarks, loc. cit., at the beginning of §11.4, and immediately preceding Lemma 11.4.3. We can use $D_{\mathbf{M}}$ and $D_{\mathbf{M}'}$ to obtain relative dimension functions in $\mathbf{M}_{(\mathfrak{M}_1 \cdot \mathfrak{M}_1')}$, $\mathbf{M}'_{(\mathfrak{M}_1 \cdot \mathfrak{M}_1')}$ respectively, as described in Lemma 11.4.2, loc. cit. Our equations (3.3.α) and (3.3.β) are now seen to imply:

First, the relative dimension of $\mathfrak{M}_1 \cdot \mathfrak{M}_1'$ is 1 for $\mathbf{M}_{(\mathfrak{M}_1 \cdot \mathfrak{M}_1')}$ and for $\mathbf{M}'_{(\mathfrak{M}_1 \cdot \mathfrak{M}_1')}$. Hence both factors are in a finite case. Besides, $C = 1$ (cf. above). Hence they are either both in a case (I_n), $n = 1, 2, \cdots$ with $1/n$ times the standard normalization, or both in case (II_1) with the standard normalization.[20]

Second, they fulfill the requirements of [6], p. 235, beginning of §4.1, on which the Theorem VI on p. 239 eod. is based.[21] Now this theorem states that the coupled factors to which it applies are dual isomorphic.[22] Again, by [5], p. 188, Lemma 11.4.1, $\mathbf{M}_{(\mathfrak{M}_1)}$, $\mathbf{M}_{(\mathfrak{M}_1 \cdot \mathfrak{M}_1')}$ are algebraically ring isomorphic and similarly $\mathbf{M}'_{(\mathfrak{M}_1')}$, $\mathbf{M}_{(\mathfrak{M}_1 \cdot \mathfrak{M}_1')}$. Thus $\mathbf{M}_{(\mathfrak{M}_1)}$, $\mathbf{M}'_{(\mathfrak{M}_1')}$ are dual isomorphic, i.e.

$$\overline{\overline{\mathbf{M}'_{(\mathfrak{M}_1')}}} = \overline{\overline{\mathbf{M}_{(\mathfrak{M}_1)}}}^c. \tag{3.3.γ}$$

Consequently,

$$\overline{\overline{\mathbf{M}'}} = \overline{\overline{\mathbf{M}'_{(\mathfrak{M}_1')}}} = \overline{\overline{\mathbf{M}_{(\mathfrak{M}_1)}}}^c = \overline{\overline{\mathbf{M}}}^c$$

as desired.

This lemma determines the direction of further analysis of our present topic. It is easy to obtain now more detailed information.

---

[20] This discrepancy is of course due to the different ways in which we defined the standard normalization in the cases (I) and (II) (cf. [5], p. 172, Theorem VIII).

[21] The theorem referred to is stated for the case (II_1) only. It holds however for the cases (I_n), $n = 1, 2, \cdots$ too, and with precisely the same proof. Our remark in footnote[1] could have been applied to [6] also.

[22] Called anti-isomorphic, loc. cit.

For this purpose we define

DEFINITION 3.3.1. *For every factor* $\mathbf{M}$ *in the cases* (I) *and* (II) *we form the number*

$$\theta = \frac{1}{C} \frac{D_{\mathbf{M}'}(\mathfrak{H})}{D_{\mathbf{M}}(\mathfrak{H})} \tag{3.3.δ}$$

with any normalization of $D_{\mathbf{M}}$, $D_{\mathbf{M}'}$ *and the corresponding* C *of* [5], *p.* 182, *Theorem* X.

The expression for θ is clearly independent of the normalization of $D_{\mathbf{M}}$, $D_{\mathbf{M}'}$.

Since $0 < C < \infty$, $0 < D_{\mathbf{M}}(\mathfrak{H})$, $D_{\mathbf{M}'}(\mathfrak{H}) \leqq \infty$, (3.3.$\delta$) characterizes θ as a well defined number with

$$0 \leqq \theta \leqq \infty \tag{3.3.ϵ}$$

except when $D_{\mathbf{M}}(\mathfrak{H}) = D_{\mathbf{M}'}(\mathfrak{H}) = \infty$ i.e. when $\mathbf{M}$, $\mathbf{M}'$ both belong to infinite cases. If this happens, we construe (3.3.δ) to mean

$$\theta = \frac{\infty}{\infty}, \tag{3.3.ζ}$$

the symbol ∞/∞ being considered as an entity different from all numbers of (3.3.ϵ).

LEMMA 3.3.2. *Let* $\mathbf{M}$ *and* θ *be as above. Then we have*

α) *If* $\mathbf{M}$, $\mathbf{M}'$ *are both in infinite cases, then*

$$\overline{\mathbf{M}'} = \overline{\mathbf{M}}^c \quad \text{and} \quad \theta = \frac{\infty}{\infty}.$$

β) *If* $\mathbf{M}$ *is in an infinite case, and* $\mathbf{M}'$ *is in a finite case, then*

$$\overline{\mathbf{M}} = (\overline{\mathbf{M}'}^c)^\infty \quad \textit{and} \quad \theta = 0 \quad \textit{so} \quad 1/\theta = \infty.$$

γ) *If* $\mathbf{M}$ *is in a finite case and* $\mathbf{M}'$ *is in an infinite case, then*

$$\overline{\mathbf{M}'} = (\overline{\mathbf{M}}^c)^\infty \quad \textit{and} \quad \theta = \infty.$$

δ) *If* $\mathbf{M}$, $\mathbf{M}'$ *are both in finite cases, then*

$$\overline{\mathbf{M}'} = (\overline{\mathbf{M}}^c)^\theta \quad \textit{or equivalently} \quad \overline{\mathbf{M}} = (\overline{\mathbf{M}'}^c)^{\frac{1}{\theta}} \quad \textit{and} \quad 0 < \theta < \infty.$$

PROOF. All assertions concerning θ are immediately by Def. 3.3.1. Let us now consider the other statements.

Ad α), β), γ). Lemma 3.3.1 states that $\overline{\mathbf{M}'}$ and $\overline{\mathbf{M}}^c$ are commensurable. Hence our α), β), γ) follow from α), β), γ) in Lemma 3.1.6 respectively.

Ad δ). The equivalence of these two equations is obvious. We consider the first one.

We choose the normalization of $D_{\mathbf{M}}$, $D_{\mathbf{M}'}$ exactly as in the proof of Lemma 3.3.1. Now (3.3.α), (3.3.β) in that proof give

$$\overline{\mathbf{M}_{(\mathfrak{M}_1)}} = \overline{\mathbf{M}}^{\frac{1}{D_{\mathbf{M}}(\mathfrak{H})}}, \overline{\mathbf{M}'_{(\mathfrak{M}'_1)}} = (\overline{\mathbf{M}'})^{\frac{1}{D_{\mathbf{M}'}(\mathfrak{H})}}$$

Then (3.3.γ) and (2.9.α) yield

$$(\overline{\mathbf{M}'})^{\frac{1}{D_{\mathbf{M}'}(\mathfrak{H})}} = \left(\overline{\mathbf{M}}^{\frac{1}{D_{\mathbf{M}}(\mathfrak{H})}}\right)^c = (\overline{\mathbf{M}}^c)^{\frac{1}{D_{\mathbf{M}}(\mathfrak{H})}}$$

or

$$\overline{\mathbf{M}'} = (\overline{\mathbf{M}}^c)^{\frac{D_{\mathbf{M}'}(\mathfrak{H})}{D_{\mathbf{M}}(\mathfrak{H})}} = (\overline{\mathbf{M}}^c)^{\theta}$$

as desired.

REMARK. The first formula of δ) holds clearly for γ) as well, while it becomes meaningless for α), β). The second formula of δ) holds clearly for β) as well, while it becomes meaningless for α), γ).

It is furthermore apparent from our proof that for δ) the exponents θ, $1/\theta$ are in the range of Theorem VII.

LEMMA 3.3.3. *Let* $\mathbf{M}$ *and* θ *be as above. Then the spatial type* $\ddot{\mathbf{M}}$ *is uniquely determined by the algebraical types,* $\overline{\mathbf{M}}$, $\overline{\mathbf{M}'}$ *and the number* θ.

PROOF. As in the introduction, §2, for each $i = 1, 2$ a Hilbert space $\mathfrak{H}_i$ and an operator ring $\mathbf{M}_i$ in $\mathfrak{H}_i$ must be given. We assume that both $\mathbf{M}_i$ are factors in case (I) or (II) and form for each $\mathbf{M}_i$ its θ_i in the sense of Def. 3.3.1. We assume further that

$$\overline{\mathbf{M}}_1 = \overline{\mathbf{M}}_2, \qquad \overline{\mathbf{M}_1'} = \overline{\mathbf{M}_2'}, \qquad \theta_1 = \theta_2 = \theta.$$

Then our task is to establish $\ddot{\mathbf{M}}_1 = \ddot{\mathbf{M}}_2$ i.e. to find an isomorphic mapping of $\mathfrak{H}_1$ on $\mathfrak{H}_2$ (cf. A) in the Introduction, §2) which carries $\mathbf{M}_1$ into $\mathbf{M}_2$.

In proving this we must distinguish several alternatives corresponding to various values of θ (the joint value of θ_i).

Observe first that the isomorphic mapping of $\mathfrak{H}_1$ on $\mathfrak{H}_2$ which carries $\mathbf{M}_1$ into $\mathbf{M}_2$ also carries $\mathbf{M}_1'$ into $\mathbf{M}_2'$. Thus we can always replace each $\mathfrak{M}_i$ by its $\mathbf{M}_i'$. This has the consequence that $\theta_1 = \theta_2 = \theta$ is replaced by $1/\theta_1 = 1/\theta_2 = 1/\theta$.

Let us now consider the alternatives for θ.

First alternative $0 < \theta < \infty$. Since we may replace θ by $1/\theta$ we can even assume that $0 < \theta \leqq 1$. For both $i = 1, 2$ the factors $\mathbf{M}_i$, $\mathbf{M}_i'$ are in finite cases. Choose the normalizations of both $D_{\mathbf{M}_i}$, $D_{\mathbf{M}_i'}$ so that $D_{\mathbf{M}_i'}(\mathfrak{H}_i) = 1$, $C_i = 1$ (we use the notation of Def. 3.3.1 for both $i = 1, 2$). So $D_{\mathbf{M}_i}(\mathfrak{H}_i) = 1/\theta_i = 1/\theta \geqq 1$. Hence 1 belongs to the range of $D_{\mathbf{M}_i}$ (cf. [5], p. 182, the discussion of the ranges of Δ, Δ_0, Δ', Δ_0' in Theorem X and preceding it). Choose accordingly a $\mathfrak{K}_i \, \eta \, \mathbf{M}_i$ with $D_{\mathbf{M}_i}(\mathfrak{K}_i) = 1$. There exists furthermore an ϵ in the range of $D_{\mathbf{M}_1}$ for which there is a $p = 1, 2, \cdots$ with $0 < \epsilon \leqq 1$, $p\epsilon = 1/\theta$.[23] Choose accordingly an $\mathfrak{L}_1 \, \eta \, \mathbf{M}_1$ with $D_{\mathbf{M}_1}(\mathfrak{L}_1) = \epsilon$, $\mathfrak{L}_1 \subsetneqq \mathfrak{K}_1$.

---

[23] The discussion of [5], p. 182, shows that the ranges of $D_{\mathbf{M}_1}$ and $D_{\mathbf{M}_1'}$ agree for the area $\leqq$ both their maxima, i.e. 1, $1/\theta$. Hence we can argue as follows:

Case (I_n), $n = 1, 2, \cdots$: The range of $D_{\mathbf{M}_1}$ is $(0, 1/n, \cdots, 1)$ now $1/\theta \geqq 1/n$ and $1/n$ is the smallest positive element of the range of $D_{\mathbf{M}_1'}$. Hence that range contains only integer multiples of $1/n$. So $1/\theta = p/n$, $p = 1, 2, \cdots$. So $\epsilon = 1/n$ meets our requirements.

Case (II_1). Choose $p = 1, 2, \cdots$ so that $1/p\theta \leqq 1$. Then $\epsilon = 1/p\theta$ fulfills $p\epsilon = 1/\theta$ and besides $\epsilon \leqq 1$, $1/\theta$. Since ϵ belongs to the range of $D_{\mathbf{M}_1}$ it belongs to the range of $D_{\mathbf{M}_1'}$ also.

Let us now use the considerations of [5], pp. 188–190, §11.4. We form, in accord with them, $\mathbf{M}_{i(\mathfrak{K}_i)}$, $\mathbf{M}'_{i(\mathfrak{K}_i)}$. (We choose $\mathfrak{H}_i$, $\mathfrak{K}_i$, $\mathfrak{H}_i$ and $\mathbf{M}_i$, $\mathbf{M}'_i$ for the $\mathfrak{H}$, $\mathfrak{M}$, $\mathfrak{M}'$ and $\mathbf{M}_1$ $\mathbf{M}'$, loc. cit.) These are coupled factors in $\mathfrak{K}_i \neq (0)$ just as $\mathbf{M}_i$, $\mathbf{M}'_i$ are in $\mathfrak{H}_i$. We have

$$\overline{\mathbf{M}'_{i(\mathfrak{K}_i)}} = \overline{\mathbf{M}'_i}.$$

Hence $\overline{\mathbf{M}'_1} = \overline{\mathbf{M}'_2}$ gives

$$\overline{\mathbf{M}'_{1(\mathfrak{K}_1)}} = \overline{\mathbf{M}'_{1(\mathfrak{K}_2)}}.$$

Besides

$$D_{\mathbf{M}_{i(\mathfrak{K}_i)}}(\mathfrak{K}_i) = D_{\mathbf{M}_i}(\mathfrak{K}_i) = 1,$$

$$D_{\mathbf{M}_{(\mathfrak{K}_i)}}(\mathfrak{K}_i) = D_{\mathbf{M}'_i}(\mathfrak{H}_i) = 1,$$

and $C_i = 1$ as before. Hence [6], p. 244, Theorem XI, applies with our $\mathfrak{K}_i$, $\mathbf{M}'_{i(\mathfrak{K}_i)}$, $\mathbf{M}_{i(\mathfrak{K}_i)}$ in place of its $\mathfrak{H}_i$, $\mathbf{M}_i$, $\mathbf{M}'_i$. There exists an isomorphic mapping $\mathfrak{J}$ of $\mathfrak{K}_1$ on $\mathfrak{K}_2$ which carries $\mathbf{M}_{1(\mathfrak{K}_1)}$ into $\mathbf{M}_{2(\mathfrak{K}_2)}$ and $\mathbf{M}'_{1,(\mathfrak{K}_1)}$ into $\mathbf{M}'_{2(\mathfrak{K}_2)}$, $\mathfrak{L}_1 \,\eta\, \mathbf{M}_1$, $\mathfrak{L}_1 \subseteq \mathfrak{K}_1$ imply $\mathfrak{L}_1 \,\eta\, \mathbf{M}_{1(\mathfrak{K}_1)}$. So $\mathfrak{J}$ carries $\mathfrak{L}_1$ into an $\mathfrak{L}_2 \,\eta\, \mathbf{M}_{2(\mathfrak{K}_2)}$. It follows that $\mathfrak{L}_2 \,\eta\, \mathbf{M}_2$ and $\mathfrak{L}_2 \subseteq \mathfrak{K}_2$. Now

$$D_{\mathbf{M}_2}(\mathfrak{L}_2) = D_{\mathbf{M}_{2(\mathfrak{K}_2)}}(\mathfrak{L}_2) = D_{\mathbf{M}_{1(\mathfrak{K}_1)}}(\mathfrak{L}_1) = D_{\mathbf{M}_1}(\mathfrak{L}_1) = \epsilon.$$

$\mathfrak{J}$ carries $\mathbf{M}_{1(\mathfrak{K}_1)(\mathfrak{L}_1)} = \mathbf{M}_{1(\mathfrak{L}_1)}$ into $\mathbf{M}_{2(\mathfrak{K}_2)(\mathfrak{L}_2)} = \mathbf{M}_{2(\mathfrak{L}_2)}$. Restrict $\mathfrak{J}$ from $\mathfrak{K}_1$ to $\mathfrak{L}_1$ and denote this restricted mapping by $\mathfrak{F}$. Then $\mathfrak{F}$ is an isomorphic mapping of $\mathfrak{L}_1$ on $\mathfrak{L}_2$ which carries $\mathbf{M}_{1(\mathfrak{L}_1)}$ into $\mathbf{M}_{2(\mathfrak{L}_2)}$.

We have $p\epsilon = 1/\theta$ and therefore for both $i = 1, 2$

$$pD_{\mathbf{M}_i}(\mathfrak{L}_i) = D_{\mathbf{M}_i}(\mathfrak{H}_i)$$

or if we introduce the projection $E_i = P_{\mathfrak{L}_i}$ then

$$D_{\mathbf{M}_i}(E_i) = (1/p)D_{\mathbf{M}_i}(1).$$

Consequently we can proceed as in the proof of Lemma 2.6.2 or 2.6.3 to obtain a set of p-order matrix units $W_{i;u,v}$, $u, v = 1, \cdots, p$ with $W_{i;1,1} = E_i$, $\sum_{u=1}^{p} W_{i;u,u} = 1$. Let the projection $F_{i;u} = W_{i;u,u}$ have range $\mathfrak{N}_{i;u}$. We note that $\mathfrak{N}_{i;1} = \mathfrak{L}_i$ and that $W_{i;u,v}$ is partially isometric with initial set $\mathfrak{N}_{i;v}$ and final set $\mathfrak{N}_{i;u}$; i.e. it is an isomorphic mapping of $\mathfrak{N}_{i;v}$ on $\mathfrak{N}_{i;u}$.

Thus $W_{2;u,1}\mathfrak{F}W_{1;1,u} = \mathfrak{F}_u$ is an isomorphic mapping of $\mathfrak{N}_{1;u}$ on $\mathfrak{N}_{2;u}$ for $u = 1, 2, \cdots, p$. For $u = 1$, $W_{i;1,1} = F_{i,1}$, $E_i = P_{\mathfrak{L}_i}$, $\mathfrak{N}_{i;1} = \mathfrak{L}_i$ and hence $\mathfrak{F}_1$ coincides with $\mathfrak{F}$. Since the $\mathfrak{N}_{i;u}$, $u = 1, \cdots, p$ are pairwise orthogonal and span together the closed linear set $\mathfrak{H}_i$ we can combine the $\mathfrak{F}_u$, $u = 1, \cdots, p$ to one isomorphic mapping $\mathfrak{F}^*$ of $\mathfrak{H}_1$ on $\mathfrak{H}_2$. On $\mathfrak{L}_1$ this $\mathfrak{F}^*$ coincides with $\mathfrak{F}_1 = \mathfrak{F}$. Hence

(3.3.η) $\qquad \mathfrak{F}^*$ carries $\mathfrak{L}_1$ into $\mathfrak{L}_2$ and in it $\mathbf{M}_{1(\mathfrak{L}_1)}$ into $\mathbf{M}_{2(\mathfrak{L}_2)}$.

By virtue of the properties of $W_{i;u,v}$ (cf. Def. 2.6.1), and owing to the definition of $\mathfrak{F}^*$ we have further

(3.3.θ) $\mathfrak{F}^*$ carries $W_{1;u,v}$ into $W_{2;u,v}$, $u, v = 1, \cdots, p$

Now $\mathbf{M}_i$ can be characterized in terms of $\mathbf{M}_{i(\mathfrak{L}_i)}$, $\mathfrak{L}_i$ and the $W_{i;u,v}$, $u, v = 1, \cdots, p$. Indeed: $A_i \epsilon \mathbf{M}_i$ is equivalent to $(W_{i;1,v}A_iW_{i;u,1})_{(\mathfrak{L}_i)} \epsilon \mathbf{M}_{i(\mathfrak{L}_i)}$ for all $u, v = 1, \cdots, p$.[24] Consequently (3.3.η) and (3.3.θ) imply that $\mathfrak{F}^*$ carries $\mathbf{M}_1$ into $\mathbf{M}_2$.

This completes the proof of the first alternative.

Second alternative: $\theta = 0$ or ∞. Since we may replace θ by $1/\theta$ it is clear that we need only consider the case $\theta = 0$. For both $i = 1, 2$ the factor $\mathbf{M}_i$ is in an infinite case and the factor $\mathbf{M}_1'$ is in a finite case. Choose the normalization of $D_{\mathbf{M}_i}$, $D_{\mathbf{M}_i'}$ so that

$$D_{\mathbf{M}_i'}(\mathfrak{H}_i) = 1, \qquad C_i = 1.$$

We may parallel our discussion of the first alternative however with $\epsilon = 1$, $p = \infty$. We form the $\mathfrak{K}_i$ accordingly. Since $\epsilon = 1$, $\mathfrak{L}_1 = \mathfrak{K}_1$. $\mathfrak{J}$ obtains as loc. cit. $\mathfrak{L}_1 = \mathfrak{K}_1$ yields $\mathfrak{L}_2 = \mathfrak{K}_2$ and $\mathfrak{F} = \mathfrak{J}$.

Let $E_i = P_{\mathfrak{L}_i}$. We proceed to obtain an infinite number of pairwise orthogonal manifolds $\mathfrak{N}_{i,1}, \mathfrak{N}_{i,2}, \cdots$ with $\mathfrak{N}_{i,1} = \mathfrak{L}_i$, $\mathfrak{N}_{i,u} \,\eta\, \mathbf{M}_i$, $D_{\mathbf{M}_i}(\mathfrak{N}_{i,u}) = 1$ for $u = 1, 2, \cdots$, and finally $\mathfrak{H}_i = [\mathfrak{N}_{i,1}, \mathfrak{N}_{i,2}, \cdots]$ (cf. [5], p. 155, Lemma 7.1.2). We obtain the $W_{i,u,v}$ as in the second part of the proof of Lemma 2.6.2.

As in the discussion of the preceding alternative, we now obtain $\mathfrak{F}_u$, $u = 1, 2, \cdots$ and finally $\mathfrak{F}^*$, and by a literal repetition of that argument, the isomorphic mapping $\mathfrak{F}^*$ of $\mathfrak{H}_1$ on $\mathfrak{H}_2$ is seen to carry $\mathbf{M}_1$ into $\mathbf{M}_2$.[25]

Thus the second alternative too is settled.

Third alternative: $\theta = \infty/\infty$. For both $i = 1, 2$ the factors $\mathbf{M}_i$, $\mathbf{M}_i'$ are in infinite cases. Choose a finite $\mathfrak{K}_1 \,\eta\, \mathbf{M}_1$, $\mathfrak{K}_1 \neq 0$ and put $P_{\mathfrak{K}_1} = E_1 \epsilon \mathbf{M}$. Under the (algebraic ring) isomorphism of $\mathbf{M}_1$, $\mathbf{M}_2$ (due to $\overline{\mathbf{M}}_1 = \overline{\mathbf{M}}_2$) this projection $E_1 \epsilon \mathbf{M}_1$ corresponds to a projection $E_2 \epsilon \mathbf{M}_2$. Let $\mathfrak{K}_2$ be the range of E_2. Then $\mathfrak{K}_2 \,\eta\, \mathbf{M}_2$, $\mathfrak{K}_2 \neq (0)$. Then the isomorphism of $\mathbf{M}_1$, $\mathbf{M}_2$ also induces one of $\mathbf{M}_{1(\mathfrak{K}_1)}$, $\mathbf{M}_{2(\mathfrak{K}_2)}$ owing to [5], p. 187, (i), in Lemma 11.3.3. Thus

(3.3.ι) $$\overline{\mathbf{M}_{1(\mathfrak{K}_1)}} = \overline{\mathbf{M}_{2(\mathfrak{K}_2)}}.$$

Also by (ii) eod., $\overline{\mathbf{M}'_{i(\mathfrak{K}_i)}} = \overline{\mathbf{M}'_i}$ hence $\overline{\mathbf{M}'_1} = \overline{\mathbf{M}'_2}$ gives

(3.3.κ) $$\overline{\mathbf{M}'_{1(\mathfrak{K}_1)}} = \overline{\mathbf{M}'_{2(\mathfrak{K}_2)}}.$$

---

[24] Put $A_{i,u,v} = W_{i:1,v}A_iW_{i:u,1}$. The forward implication is obvious; the reverse follows from the formula

$$A = \sum_{u,v=1}^{p} W_{i;v,1}\, A_{i,u,v}\, W_{i;1,u},$$

which is easily verified.

[25] Footnote[24] applies to this case too. The convergence of $\sum_{u,v=1}^{\infty}$ in footnote[24] is readily seen to be a consequence of the convergence of the sum $\sum_{u=1}^{\infty} W_{i:u,u}$. And in connection with the latter convergence, the remarks made in the proof of (iii) in Lemma 2.6.1 apply again.

Now choose the normalization of $D_{\mathbf{M}_i}$, $D_{\mathbf{M}'_i}$ so that $D_{\mathbf{M}_i}(\mathfrak{K}_i) = 1$, $C_i = 1$. Automatically $D_{\mathbf{M}'_i}(\mathfrak{H}_i) = \infty$. We may then write

$$D_{\mathbf{M}_{i(\mathfrak{K}_i)}}(\mathfrak{K}_i) = D_{\mathbf{M}_i}(\mathfrak{K}_i) = 1$$

$$D_{\mathbf{M}'_{i(\mathfrak{K}_i)}}(\mathfrak{K}_i) = D_{\mathbf{M}'_i}(\mathfrak{H}_i) = \infty$$

and $C_i = 1$. Apply Def. 3.3.1 to $\mathfrak{K}_i$, $\mathbf{M}_{i(\mathfrak{K}_i)}$, $\mathbf{M}'_{i(\mathfrak{K}_i)}$ in place of $\mathfrak{H}$, $\mathbf{M}$, $\mathbf{M}'$ and denote the resulting number by θ_i^0. The above equations give

$$\text{(3.3.}\lambda\text{)} \qquad \theta_1^0 = \theta_2^0 = \infty.$$

(3.3.ι), (3.3.κ) and (3.3.λ) permit us to apply the second alternative, which we have already disposed of, to $\mathfrak{K}_i$, $\mathbf{M}_{i(\mathfrak{K}_i)}$, $\mathbf{M}'_{i(\mathfrak{K}_i)}$. There exists an isomorphic mapping $\mathfrak{J}$ of $\mathfrak{K}_1$ on $\mathfrak{K}_2$ which carries $\mathbf{M}_{1(\mathfrak{K}_1)}$ into $\mathbf{M}_{2(\mathfrak{K}_2)}$ and $\mathbf{M}'_{1(\mathfrak{K}_1)}$ into $\mathbf{M}'_{2(\mathfrak{K}_2)}$.

We can continue from here on exactly as in the discussion of the second alternative, i.e. paralleling the considerations of the first alternative. We have the $\mathfrak{K}_i$ already. Again we put $\epsilon = 1$, $p = \infty$. Since $\epsilon = 1$, $\mathfrak{L}_1 = \mathfrak{K}_1$. We already have $\mathfrak{J}$. $\mathfrak{L}_1 = \mathfrak{K}_1$ gives $\mathfrak{L}_2 = \mathfrak{K}_2$ and $\mathfrak{F} = \mathfrak{J}$. We have $E_i = P_{\mathfrak{K}_i} = P_{\mathfrak{L}_i}$. The $F_{i,u}$, $u = 1, 2, \cdots$ and $W_{i,u,v}$, $u, v = 1, 2, \cdots$ obtain as loc. cit. So do the $\mathfrak{F}_u$, $u = 1, 2, \cdots$ and finally $\mathfrak{F}^*$; and exactly as loc. cit., the isomorphic mapping $\mathfrak{F}^*$ of $\mathfrak{H}_1$ on $\mathfrak{H}_2$ is seen to carry $\mathbf{M}_1$ into $\mathbf{M}_2$.

Thus the third alternative too is settled and the entire proof is completed.

We settle now the question of existence.

LEMMA 3.3.4. *A factor* $\mathbf{M}$ *in the cases* (I) *or* (II) *with given algebraical types* $\overline{\mathbf{M}}$, $\overline{\mathbf{M}}'$ *and given number* θ (*cf. Def.* 3.3.1) *exists if and only if these fulfill the alternative conditions* α)–δ) *in Lemma* 3.3.2.

PROOF. Necessity: This is merely the assertion of Lemma 3.3.2.

Sufficiency: We shall continue to denote the two prescribed algebraical types by $\overline{\mathbf{M}}$, $\overline{\mathbf{M}'}$ although we have not yet identified them as belonging in this way to any coupled factors $\mathbf{M}$, $\mathbf{M}'$. (In fact, this is what we propose to prove.) The relations α)–δ) in Lemma 3.3.2, which we assumed, imply for the genera

$$\text{(3.3.}\mu\text{)} \qquad \overline{\overline{\mathbf{M}'}} = \overline{\overline{\mathbf{M}}}^c.$$

(This cannot be deduced from Lemma 3.3.1, since $\mathbf{M}$, $\mathbf{M}'$ are fictitions, cf. above.) The genus $\overline{\overline{\mathbf{M}}}$ contains a (unique) infinite type $\overline{\mathbf{R}}$ (cf. Theorem IX)—$\mathbf{R}$ being an actual factor in a Hilbert space $\mathfrak{K}$.

Consider now the factor $\mathbf{R}'$ in $\mathfrak{K}$. We repeat the construction in §2.4, immediately preceding Theorem IV, with our $\mathfrak{K}$, $\mathbf{R}'$ in place of its $\mathfrak{H}$, $\mathbf{M}$ and with its $p = \infty$. Thus using the notations of §2.4, loc. cit., our $\mathbf{R}'$ is a factor in $\mathfrak{K} = \mathfrak{H} = \mathfrak{H}_2$ and with the ($p = \infty$ dimensional) Hilbert space $\mathfrak{H}_1$, we obtain the Hilbert space $\mathfrak{H}_1 \otimes \mathfrak{H}_2$ and in it the ring $\mathbf{M}_1$ of all operators $\langle A'_{t,s} \rangle$ with all $A'_{t,s} \in \mathbf{R}'$ (cf. the similar argument before Theorem IV in §2.4). Put $\mathbf{R}_1 = \mathbf{M}'_1$ i.e. $\mathbf{M}_1 = \mathbf{R}'_1$. It is obvious by the usual matrix computations, that $\mathbf{R}_1 = \mathbf{M}'_1$ is the set of all operators $\langle A\delta_{t,s} \rangle$, $A \in \mathbf{R}$. Hence $\overline{\mathbf{R}}_1 = \overline{\mathbf{R}}$. On the other hand, the discussions of §2.4 show that $\overline{\mathbf{R}'_1} = \overline{\mathbf{M}}_1 = \overline{\mathbf{R}'}^{\infty}$. Thus $\overline{\mathbf{R}'_1}$ is in an infinite case (by Lemma 2.7.1).

Now $\bar{\mathbf{R}}_1 = \bar{\mathbf{R}}$ permits us to replace $\mathfrak{K}$, $\mathbf{R}$ by $\mathfrak{H}_1 \otimes \mathfrak{H}_2$, $\mathbf{R}_1$. So we have shown: It is permissible to assume that $\overline{\mathbf{R}'}$ is in an infinite case. We do this, but we continue using the original notations $\mathfrak{K}$, $\mathbf{R}$.

By Lemma 3.3.1, $\overline{\overline{\mathbf{R}'}} = \bar{\bar{\mathbf{R}}}^c$. Now $\bar{\bar{\mathbf{R}}} = \bar{\bar{\mathbf{M}}}$; hence $\overline{\overline{\mathbf{R}'}} = \bar{\bar{\mathbf{M}}}^c$. By (3.3.$\mu$) this means $\overline{\overline{\mathbf{R}'}} = \bar{\bar{\mathbf{M}}}'$. Since $\bar{\mathbf{R}}$, $\overline{\mathbf{R}'}$ are both infinite, we have (cf. Theorem IX),

(3.3.ν) $\{\bar{\mathbf{R}}, \overline{\mathbf{R}'}$ are the (unique) infinite types of the genera, $\bar{\bar{\mathbf{M}}}$, $\overline{\overline{\mathbf{M}'}}$ respectively.

Choose an $f_0 \neq 0$ in $\mathfrak{K}$ and form the closed linear sets $\mathfrak{M}_{f_0}^{\mathbf{R}'} \,\eta\, \mathbf{R}$, $\mathfrak{M}_{f_0}^{\mathbf{R}} \,\eta\, \mathbf{R}'$. (cf. [5], p. 143, Def. 5.1.1). We may assume that $\mathfrak{M}_{f_0}^{\mathbf{R}'}$ is finite dimensional. (Choose a finite dimensional $\mathfrak{M} \,\eta\, \mathbf{R}$, $\mathfrak{M} \neq (0)$ and then $f_0 \in \mathfrak{M}$. Clearly $\mathfrak{M}_{f_0}^{\mathbf{R}'} \subseteq \mathfrak{M}$.) In this case, $\mathfrak{M}_{f_0}^{\mathbf{R}}$ is automatically finite dimensional. (Use [5], p. 182, Theorem X, for this and for the next step.) Hence both $D_{\mathbf{R}}(\mathfrak{M}_{f_0}^{\mathbf{R}'})$, $D_{\mathbf{R}'}(\mathfrak{M}_{f_0}^{\mathbf{R}})$ are >0, $<\infty$ and $C = D_{\mathbf{R}'}(\mathfrak{M}_{f_0}^{\mathbf{R}})/D_{\mathbf{R}}(\mathfrak{M}_{f_0}^{\mathbf{R}'})$. Now normalize $D_{\mathbf{R}}$, $D_{\mathbf{R}'}$ so that $D_{\mathbf{R}}(\mathfrak{M}_{f_0}^{\mathbf{R}'}) = D_{\mathbf{R}'}(\mathfrak{M}_{f_0}^{\mathbf{R}})$. The above formula for C gives

(3.3.o) $$C = 1.$$

Write $\mathfrak{M}_1 = \mathfrak{M}_{f_0}^{\mathbf{R}'}$, $\mathfrak{M}_1' = \mathfrak{M}_{f_0}^{\mathbf{R}}$ then

(3.3.ξ) $$D_{\mathbf{R}}(\mathfrak{M}_1) = D_{\mathbf{R}'}(\mathfrak{M}_1').$$

We shall now again make use of the considerations of [5], pp. 188–190, §11.4. In accord with them we form $\mathbf{R}_{(\mathfrak{M}_1\cdot\mathfrak{M}_1')}$, $\mathbf{R}'_{(\mathfrak{M}_1\cdot\mathfrak{M}_1')}$. These are coupled factors in $\mathfrak{M}_1\cdot\mathfrak{M}_1' \neq (0)$ just as $\mathbf{R}$, $\mathbf{R}'$ are in $\mathfrak{H}$, and (3.3.ξ) gives

$$D_{\mathbf{R}_{(\mathfrak{M}_1\cdot\mathfrak{M}_1')}}(\mathfrak{M}_1\cdot\mathfrak{M}_1') = D_{\mathbf{R}'_{(\mathfrak{M}_1\cdot\mathfrak{M}_1')}}(\mathfrak{M}_1\cdot\mathfrak{M}_1').$$

This, together with (3.3.o) and a normalization, shows that the requirements of [6], p. 239, Theorem VI, are fulfilled.[21] Hence $R_{(\mathfrak{M}_1\cdot\mathfrak{M}_1')}$, $R'_{(\mathfrak{M}_1\cdot\mathfrak{M}_1')}$ are dual isomorphic,[22] i.e.

$$\overline{\mathbf{R}'_{(\mathfrak{M}_1\cdot\mathfrak{M}_1')}} = \overline{\mathbf{R}_{(\mathfrak{M}_1\cdot\mathfrak{M}_1')}}^{\,c}$$

Now $\overline{\mathbf{R}_{(\mathfrak{M}_1)}} = \overline{\mathbf{R}_{(\mathfrak{M}_1\cdot\mathfrak{M}_1')}}$, $\overline{\mathbf{R}'_{(\mathfrak{M}_1')}} = \overline{\mathbf{R}'_{(\mathfrak{M}_1\cdot\mathfrak{M}_1')}}$
Consequently

(3.3.π) $$\overline{\mathbf{R}'_{(\mathfrak{M}_1')}} = \overline{\mathbf{R}_{(\mathfrak{M}_1)}}^{\,c}.$$

Since $\bar{\bar{\mathbf{R}}} = \bar{\bar{\mathbf{M}}}$, $\bar{\mathbf{R}}$, $\bar{\mathbf{M}}$ are commensurable. $\bar{\mathbf{R}}$ is in an infinite case. Hence α), β) in Lemma 3.1.6 and Lemma 3.1.4 give together that $\bar{\mathbf{M}}$ is a divisor of $\bar{\mathbf{R}}$. Hence by Def. 3.1.1,

(3.3.ρ) $$\bar{\mathbf{M}} = \bar{\mathbf{R}}_{(\mathfrak{L})} \qquad \textit{for a suitable } \mathfrak{L} \,\eta\, \mathbf{R}.$$

If $\mathfrak{L}$ is of finite relative dimension, we may assume by [5], Theorem X that $\mathfrak{L} = \mathfrak{M}_{f_0}^{\mathbf{R}'}$ for some f_0. Similarly

(3.3.σ) $$\overline{\mathbf{M}'} = \overline{\mathbf{R}'_{(\mathfrak{L}')}} \qquad \textit{for a suitable } \mathfrak{L}' \,\eta\, R',\ \mathfrak{L}' = \mathfrak{M}_{g_0}^{\mathbf{R}},\ \textit{when finite.}$$

From here on, it is again convenient to distinguish several alternatives, corresponding to various values of θ.

First alternative: $0 < \theta < \infty$. We have the case δ) in Lemma 3.3.2. Hence $\overline{\overline{\mathbf{M}'}} = (\overline{\overline{\mathbf{M}}}^c)^\theta$ or by (3.3.ρ) and (3.3.σ)

$$\overline{\overline{\mathbf{R}'_{(\mathfrak{L}')}}} = (\overline{\overline{\mathbf{R}_{(\mathfrak{L})}}}^c)^\theta$$

Now (3.3.π) and (i) in Lemma 3.2.2, gives

$$\overline{\overline{\mathbf{R}'_{(\mathfrak{L}')}}} = (\overline{\overline{\mathbf{R}_{\mathfrak{L}}}}^c)^\beta$$

where $\beta = D_{\mathbf{R}'}(\mathfrak{L}')/D_{\mathbf{R}}(\mathfrak{L})$. Hence (ii) in Lemma 3.2.2 give

$$\alpha = \beta/\theta = (1/\theta)(D_{\mathbf{R}'}(\mathfrak{L}')/D_{\mathbf{R}}(\mathfrak{L})) \in \mathfrak{G}(\overline{\overline{\mathbf{R}}}) = \mathfrak{G}(\overline{\overline{\mathbf{R}_{(\mathfrak{L})}}})$$

The interchange of $\overline{\overline{\mathbf{M}}}$ and $\overline{\overline{\mathbf{M}'}}$ and with them of θ and $1/\theta$ as well as $\mathbf{R}$, $\mathfrak{L}$ and $\mathbf{R}'$, $\mathfrak{L}'$ replaces α by $1/\alpha$. So we may assume $\alpha \leqq 1$. Hence $\alpha \in \mathfrak{G}(\overline{\overline{\mathbf{R}_{(\mathfrak{L})}}})$ secures the constructability of $\overline{\overline{\mathbf{R}_{(\mathfrak{L})}}}^\alpha$ in the original sense of §2.8. Thus there exists an $\mathfrak{L}'' \subset \mathfrak{L}$ with $\mathfrak{L}'' \,\eta\, \mathbf{R}_{(\mathfrak{L})}$ and

$$\alpha = \frac{D_{\mathbf{R}(\mathfrak{L}'')}(\mathfrak{L}'')}{D_{\mathbf{R}(\mathfrak{L})}(\mathfrak{L})} = \frac{D_{\mathbf{R}}(\mathfrak{L}'')}{D_{\mathbf{R}}(\mathfrak{L})}$$

($\mathfrak{L}'' \,\eta\, \mathbf{R}_{(\mathfrak{L})}$ implies $\mathfrak{L}'' \,\eta\, \mathbf{R}$.) $\alpha \in \mathfrak{G}(\overline{\overline{\mathbf{R}_{(\mathfrak{L})}}})$ gives

$$\overline{\overline{\mathbf{R}_{(\mathfrak{L}'')}}} = \overline{\overline{\mathbf{R}_{(\mathfrak{L})}}}^\alpha = \overline{\overline{\mathbf{R}_{(\mathfrak{L})}}} = \overline{\overline{\mathbf{M}}}.$$

Thus we may replace in all our considerations $\mathfrak{L}$ by $\mathfrak{L}''$. This carries $\alpha = 1$. Writing again $\mathfrak{L}$ for $\mathfrak{L}''$, we see: We may assume that $\alpha = 1$ i.e. that

$$\theta = \frac{D_{\mathbf{R}'}(\mathfrak{L}')}{D_{\mathbf{R}}(\mathfrak{L})}.$$

Now consider the coupled factors $\mathbf{R}_{(\mathfrak{L}\cdot\mathfrak{L}')}$, $\mathbf{R}'_{(\mathfrak{L}\cdot\mathfrak{L}')}$ in $\mathfrak{L}\cdot\mathfrak{L}' \neq (0)$. Combining our results, we see that we have

$$\overline{\overline{\mathbf{R}_{(\mathfrak{L}\cdot\mathfrak{L}')}}} = \overline{\overline{\mathbf{R}_{(\mathfrak{L})}}} = \overline{\overline{\mathbf{M}}}.$$

$$\overline{\overline{\mathbf{R}'_{(\mathfrak{L}\cdot\mathfrak{L})}}} = \overline{\overline{\mathbf{R}'_{(\mathfrak{L}')}}} = \overline{\overline{\mathbf{M}'}}.$$

Thus the spatial type $\dddot{\mathbf{R}}_{(\mathfrak{L}\cdot\mathfrak{L}')}$ meets all our requirements.

Second alternative: $\theta = 0$ or $\theta = \infty$. Since we may interchange $\overline{\overline{\mathbf{M}}}$ and $\overline{\overline{\mathbf{M}'}}$ and with them θ and $1/\theta$ we can even assume that $\theta = 0$. Then we have the case β) in Lemma 3.3.2.

Thus $\overline{\overline{\mathbf{M}}}$ is in an infinite case. Hence (3.3.ν) gives $\overline{\overline{\mathbf{R}}} = \overline{\overline{\mathbf{M}}}$. Now consider the coupled factors $\mathbf{R}_{(\mathfrak{L}_1)}$, $\mathbf{R}'_{(\mathfrak{L}')}$ in $\mathfrak{L}' \neq (0)$. Combining our results we see that we have

$$\overline{\overline{\mathbf{R}_{(\mathfrak{L}')}}} = \overline{\overline{\mathbf{R}}} = \overline{\overline{\mathbf{M}}}.$$

$$\overline{\overline{\mathbf{R}'_{(\mathfrak{L}')}}} = \overline{\overline{\mathbf{M}'}}$$

$$D_{\mathbf{R}_{(\mathfrak{L}')}}(\mathfrak{L}) = D_{\mathbf{R}}(\mathfrak{K}) = \infty.$$

$$D_{\mathbf{R}'_{(\mathfrak{L}')}}(\mathfrak{L}') = D_{\mathbf{R}'}(\mathfrak{L}') < \infty.$$

Thus the spatial type $\ddot{\mathbf{R}}_{(\mathfrak{L}')}$ meets all our requirements.

Third alternative: $\theta = \infty/\infty$. We have the case α) in Lemma 3.3.2. Hence $\overline{\mathbf{M}}$, $\overline{\mathbf{M}'}$ are both in infinite cases, and so (3.3.ν) gives

$$\bar{\mathbf{R}} = \overline{\mathbf{M}}$$
$$\overline{\mathbf{R}'} = \overline{\mathbf{M}'}.$$

Also

$$D_{\mathbf{R}}(\mathfrak{K}) = D_{\mathbf{R}'}(\mathfrak{K}) = \infty.$$

Thus the spatial type $\ddot{\mathbf{R}}$ meets all our requirements.

This completes the proof.

Summing up the results of Lemmas 3.3.3 and 3.3.4,

THEOREM X. *Consider factors* **M** *in the cases* (I) *and* (II). *Then the spatial type* $\ddot{\mathbf{M}}$ *is uniquely determined by the algebraical types* $\overline{\mathbf{M}}$, $\overline{\mathbf{M}'}$ *and the number* θ (*cf. Def.* 3.3.1).

$\ddot{\mathbf{M}}$ *exists if and only if* $\overline{\mathbf{M}}$, $\overline{\mathbf{M}'}$ *and* θ *fulfill the alternative conditions* α) $\cdots$ δ) *in Lemma* 3.3.2.

REMARK. In some cases, two of $\overline{\mathbf{M}}$, $\overline{\mathbf{M}'}$ and θ suffice to determine the third one and thus $\ddot{\mathbf{M}}$. Specifically,

α) If $\overline{\mathbf{M}}$ is in a finite case, then $\overline{\mathbf{M}}$ and θ determine $\overline{\mathbf{M}'}$ by γ), δ) in Lemma 3.3.2.

β) If $\overline{\mathbf{M}'}$ is in a finite case, then $\overline{\mathbf{M}'}$ and θ determine $\overline{\mathbf{M}}$ by β), δ) in Lemma 3.3.2.

γ) If either $\overline{\mathbf{M}}$ or $\overline{\mathbf{M}'}$ is in an infinite case, then $\overline{\mathbf{M}}$, $\overline{\mathbf{M}'}$ determine θ by α), β) or γ) in Lemma 3.3.2.

δ) If $\overline{\mathbf{M}}$, $\overline{\mathbf{M}'}$ are both in finite cases, then $\overline{\mathbf{M}}$, $\overline{\mathbf{M}'}$ determine θ precisely up to a factor belonging to the group $\mathfrak{G}(\bar{\bar{\mathbf{M}}})$ owing to δ) in Lemma 3.3.2 and to (2.10.α) and Lemma 3.2.1. Consequently they determine θ outright if and only if $\mathfrak{G}(\bar{\bar{\mathbf{M}}}) = (1)$.[26]

Thus the only remaining objects for our inquiries in the cases (I) and (II) are the algebraical types $\overline{\mathbf{M}}$. Theorems IX and X (or Lemmas 3.2.1 and 3.2.2 in place of the latter) show that the knowledge of these is equivalent to the knowledge of their genera $\bar{\bar{\mathbf{M}}}$. The case (I) is completely clear; hence the above concepts must be studied in the case (II). And we can analyze a genus $\bar{\bar{\mathbf{M}}}$ in case (II) by analyzing any one of its types $\overline{\mathbf{M}}$ in a finite case. Consequently we need only be concerned with (algebraical) types $\overline{\mathbf{M}}$ in case (II_1).

CHAPTER IV. APPROXIMATE FINITENESS

§4.1 In accordance with the above we begin the systematic investigation of the algebraical types $\overline{\mathbf{M}}$ in case (II_1). Throughout this chapter **M** will be a factor in case (II_1) in Hilbert space $\mathfrak{H}$. We use $D_{\mathbf{M}}(E)$, (but now in the standard

[26] Hence θ is never needed in the cases (I) by Lemma 2.10.4 The situation is different in the cases (II). Cf. Theorem XV and the remark at the end of §5.6.

normalization) $Tr_{\mathbf{M}}(A)$, $[[A]]$ as in §1.2. We shall also consider other factors in $\mathfrak{H}$ but they will be denoted by different letters. All our considerations however will take place in $\mathfrak{H}$.

We begin by stating some familiar facts about factors in the case (I_n), $n = 1, 2, \cdots$.

LEMMA 4.1.1. *Let* $\mathbf{N}$ *be a factor in a case* (I_p), $p = 1, 2, \cdots$. *Then there exists a one-to-one correspondence between the two following sets of entities,*

(α) *The (algebraical ring) isomorphisms* $\mathcal{K}$ *between* $\mathbf{N}$ *and the system of all complex p-order matrices*

$$a = \{a_{t,s}\}, \qquad (t, s = 1, \cdots, p)^{27}$$

(β) *The matrix bases of* $\mathbf{N}$, *i.e. the systems* $W_{u,v}$, $(u, v = 1, \cdots, p)$ *of p order matrix units, that are contained in* $\mathbf{N}$ *and have*

$$E_0 = \textstyle\sum_{u=1}^{p} W_{u,u} = 1. \tag{4.1.α}$$

The correspondence is this:

(i) $\mathcal{K}$ *is given:*

$$W_{u,v} = \mathcal{K}^{-1}\{\delta_{t,v}\delta_{s,u}\}$$

(ii) *The* $W_{u,v}$ *are given:*

$$\mathcal{K}A = a = \{a_{t,s}\}$$

is equivalent to

$$A = \textstyle\sum_{u=1}^{p} \sum_{v=1}^{p} a_{v,u} W_{u,v}$$

PROOF.[28] (i) leads from (α) to (β). The conditions of Def. 2.6.1 as well as (4.1.α) are verified by direct matrix computation.

(ii) leads from (β) to (α): The rules of matrix computation are verified from Def. 2.6.1 and (4.1.α).

(i) implies (ii): Immediate by matrix computation.

(ii) implies (i). Put $a_{u,v} = \delta_{t,v}\delta_{s,u}$ in (ii).

REMARK 1. It is clear by (ii) that the ring $\mathbf{N}$ is generated by the $W_{u,v}$, $u, v = 1, \cdots, p$,

$$\mathbf{N} = \mathbf{R}(W_{u,v}\,;\, u, v = 1, \cdots, p).$$

REMARK 2. The two sets (α), (β) are of course not empty. This follows in numerous ways from past results; the equivalent was even stated explicitly in the Corollary to Lemma 2.10.3.

LEMMA 4.1.2. *Let* $\mathbf{N}$, $\mathbf{O}$ *be two factors in the cases* (I_p), (I_q) *respectively,* $p, q, = 1, 2, \cdots$. *Assume that*

$$\mathbf{N} \subseteq \mathbf{O}.$$

[27] The (complex numerical) matrices which we consider now should be viewed as endowed with the same operations and defined in the same way as the (operator) matrices of B_p) in §2.4.

[28] Given only for the sake of completeness.

Then we have

(i) *p is a divisor of q: $q = pr$, $r = 1, 2, \cdots$.*

(ii) *To every matrix basis, $W_{u,v}$, $u, v = 1, \cdots, p$ of* $\mathbf{N}$ *there exists a matrix base $X_{o,w}$, $o, w = 1, \cdots, q$ of* $\mathbf{O}$ *such that*

$$W_{u,v} = \sum_{i=1}^{r} X_{(u-1)r+i,(v-1)r+i}$$

(iii) *If we pass from the matrix bases $W_{u,v}$ and $X_{o,w}$ in* (ii) *to the isomorphism $\mathcal{K}$ and $\mathcal{L}$ which correspond to them by Lemma* 4.1.1, *then the correlation of* (ii) *is equivalent to this*:

For $A \in \mathbf{N} \subseteq \mathbf{O}$, both $\mathcal{K}A = a = \{a_{t,s}\}$, $t, s = 1, \cdots, p$ and $\mathcal{L}A = \bar{a} = \{\bar{a}_{k,l}\}$ $k, l = 1, \cdots, q$ are defined and

$$\bar{a}_{k,l} = a_{t,s}\delta_{i,j} \quad \text{for} \quad k = (t-1)r + i, \qquad l = (s-1)r + j.$$

PROOF. Ad (i). $D_{\mathbf{O}}(E)$ will do as $D_{\mathbf{N}}(E)$ for the projections $E \in \mathbf{N} \subseteq \mathbf{O}$ in the sense of [5], p. 165, Def. 8.2.1[29] and standard normalization. And since the dimension is uniquely characterized by those properties, therefore $D_{\mathbf{O}}(E) = D_{\mathbf{N}}(E)$ for these E.

Hence the total range of $D_{\mathbf{O}}(E)$ contains the total range of $D_{\mathbf{N}}(E)$ i.e. the set $(0, 1/q, \cdots, 1)$ contains the set $(0, 1/p, \cdots, 1)$. Consequently, p is a divisor of q.

Ad (ii). Consider a matrix base $W_{u,v}$ of $\mathbf{N}$ and an arbitrary matrix base $X'_{o,w}$ of $\mathbf{O}$. Put

$$W'_{u,v} = \sum_{i=1}^{r} X_{(u-1)r+i,(v-1)r+i}\,. \tag{4.1.β}$$

Then one verifies by explicit computation (using Def. 2.6.1) that the $W'_{u,v}$ also form a system of p-order matrix units. Also

$$\sum_{u=1}^{p} W'_{u,u} = \sum_{o=1}^{q} X'_{o,o} = 1.$$

Since $D_{\mathbf{O}}(W'_{u,u}) = D_{\mathbf{O}}(W'_{1,1})$ we have $D_{\mathbf{O}}(W'_{1,1}) = 1/p$ and similarly $D_{\mathbf{O}}(W_{1,1}) = 1/p$. Therefore there exists a partially isometric U in $\mathbf{O}$ with the initial projection $W'_{1,1}$ and the final projection $W_{1,1}$. Now put

$$V = \sum_{u=1}^{p} W_{u,1} U W'_{1,u}\,.$$

Clearly $V \in \mathbf{O}$ and one verifies by direct computation

(1) $\qquad V^*V = VV^* = 1 \qquad$ i.e. V is unitary.

(2) $\qquad VW'_{u,v} = W_{u,v}V \qquad$ i.e. $W_{u,v} = VW'_{u,v}V^{-1}$

Hence replacement of $X'_{u,v}$ by $VX'_{u,v}V^{-1} = X_{u,v}$ will leave all its properties intact, but replace $W'_{u,v}$ by $W_{u,v} = VW'_{u,v}V^{-1}$. And now (4.1.$\beta$) yields the desired equation of (ii).

Ad (iii). Apply (ii) in Lemma 4.1.1 to $\mathcal{K}$ (with $W_{u,v}$) and $\mathcal{L}$ (with $X_{o,w}$). Then the equation of (ii) goes over into the desired relation of (iii).

---

[29] Consider also p. 170, Lemma 8.3.5, loc. cit. We are using the fact that $\mathbf{O}$ belongs to a finite case.

We now define

DEFINITION 4.1.1. **M** *is approximately finite of type* $[p_1, p_2, \cdots]$ *if this is true. There exists a sequence of factors* $\mathbf{N}_1, \mathbf{N}_2, \cdots$ *with the following properties:*

(i) $\mathbf{N}_n$ *is in case* (I_{p_n})

(ii) $\mathbf{N}_1 \subset \mathbf{N}_2 \subset \cdots$.

(iii) *The ring* **M** *is generated by the* $\mathbf{N}_1, \mathbf{N}_2, \cdots$.

$$\mathbf{M} = \mathbf{R}(\mathbf{N}_n ; n = 1, 2, \cdots)$$

We have immediately:

LEMMA 4.1.3. *Under the conditions of Def.* 4.1.1, *we have necessarily*

(i) p_n *is a divisor of* p_{n+1}, $p_{n+1} = p_n r_n$ $(r_n = 1, 2, \cdots)$.

(ii) $p_n \to \infty$ *for* $n \to \infty$ *i.e. infinitely many* $r_n > 1$.

PROOF. Ad (i). As $\mathbf{N}_n \subseteq \mathbf{N}_{m+1}$ this follows from (i) in Lemma 4.1.2.

Ad (ii). Owing to (i), $p_1 \leqq p_2 \leqq \cdots$ so failure of (ii) implies $p_m = p_{m+1} = p_{m+2} = \cdots$ for some m. Hence equally $\mathbf{N}_m = \mathbf{N}_{m+1} = \mathbf{N}_{m+2} = \cdots$[30] and so (iii) in Def. 4.1.1 gives

$$\mathbf{M} = \mathbf{R}(\mathbf{N}_n, n = 1, 2, \cdots) = \mathbf{N}_m.$$

But this is impossible since **M** is in the case (II_1) and $\mathbf{N}_m$ in the case (I_{p_m}).

For every sequence $[p_1, p_2, \cdots]$ which meets these requirements there exists an **M** which is approximately finite of that type. But the concept introduced by Def. 4.1.1 is a preliminary one and we prefer to prove this existence only in conjunction with the final reformulation of the concept of approximate finiteness in Theorems XII and XIV. We now proceed in a different direction.

LEMMA 4.1.4. *For any sequence of rings* $\mathbf{N}_1, \mathbf{N}_2, \cdots$ *which fulfills conditions* (ii), (iii) *in Def.* 4.1.1, *denote by* **S** *the set theoretical sum of that sequence. Then we have*

(i) **S** *is an algebra.*

(ii) **M** *is the set of elements of* **B** *which are limits of a convergent sequence from* **S**.

PROOF. Ad (i): Every $\mathbf{N}_n$ is a ring; hence an algebra. Also $\mathbf{N}_1 \subset \mathbf{N}_2 \subset \cdots$. Hence the sum **S** is an algebra also.

Ad (ii): Denote the set of limits of all metrically convergent sequences from **S** by $\mathbf{S}_1$. We must prove that $\mathbf{S}_1 = \mathbf{M}$.

Now

$$\mathbf{M} = \mathbf{R}(\mathbf{N}_n, n = 1, 2, \cdots) = \mathbf{R}(\mathbf{S})$$

and $\mathbf{S} \subseteq \mathbf{S}_1 \subseteq \mathbf{M}$. Hence $\mathbf{S}_1 = \mathbf{M}$ is established if we can show that $\mathbf{S}_1$ is a ring.

S is an algebra and αA, A^* (as one-variable functions), $A + B$ (as a two-variable function), and AB (as a one-variable function with respect to either variable) are continuous for metric convergence. (This follows from the evaluations of [6], p. 242, property II°.) Hence $\mathbf{S}_1$ is also an algebra. Therefore

[30] For $l \geqq m$, $\mathbf{N}_m$ and $\mathbf{N}_l$ contain the same number of linearly independent elements: $p_l^2 = p_m^2$. As $\mathbf{N}_m \subseteq \mathbf{N}_l$ this implies $\mathbf{N}_m = \mathbf{N}_l$.

its closure with respect to metric convergence implies its closure in the weak topology by Theorem I. Consequently, $\mathbf{S}_1$ is a ring.

This completes the proof.

We can now prove

THEOREM XI. *The (algebraical) type* $\overline{\mathbf{M}}$ *of an* $\mathbf{M}$ *which is approximately finite of type* $[p_1, p_2, \cdots]$ *depends on the sequence* $[p_1, p_2, \cdots]$ *alone and not on the particular choice of* $\mathbf{M}$.

(Provided such an $\mathfrak{M}$ exists at all. Cf. above.)

PROOF. Consider two such $\mathbf{M}^{(h)}$, $h = 1, 2$, and form their $\mathbf{N}_1^{(h)}, \mathbf{N}_2^{(h)}, \cdots$ by Def. 4.1.1 and their $\mathbf{S}^{(h)}$ by Lemma 4.1.4. We wish to prove that $\mathbf{M}^{(1)}$ and $\mathbf{M}^{(2)}$ are algebraically isomorphic.

We shall choose by induction for every $n = 1, 2, \cdots$ an (algebraical) isomorphism $\mathcal{K}_n^{(h)}$ between $\mathbf{N}_n^{(h)}$ and the system of all (complex) p_n-order matrices in the sense of (α) in Lemma 4.1.1. For $n = 1$ we choose a $\mathcal{K}_1^{(h)}$ which meets this requirement (cf. Remark 2 after Lemma 4.1.1). If for any $n = 2, 3, \cdots$, $\mathcal{K}_{n-1}^{(h)}$ has already been chosen, then we choose $\mathcal{K}_n^{(h)}$ in the following way: Apply Lemma 4.1.2 to $\mathbf{N}_{(n-1)}^{(h)}, \mathbf{N}_n^{(h)}, \mathcal{K}_{n-1}^{(h)}$ in place of its $\mathbf{N}$, $\mathbf{O}$, $\mathcal{K}$ (with $\mathbf{M}^{(h)}$ in place of $\mathbf{M}$) and put $\mathcal{K}_n^{(h)} = \mathcal{L}$. Thus $\mathcal{K}_{n-1}^{(h)}$ and $\mathcal{K}_n^{(h)}$ are related to each other as $\mathcal{K}$ and $\mathcal{L}$ in (iii) in Lemma 4.1.2.

Now form the (algebraical) isomorphism $\mathfrak{J}_n = (\mathcal{K}_n^{(2)})^{-1}\mathcal{K}_n^{(1)}$, of $\mathbf{N}_n^{(1)}$ and $\mathbf{N}_n^{(2)}$. In view of the relationship given in the last sentence of the preceding paragraph, $\mathfrak{J}_n$ is a part of $\mathfrak{J}_{n+1}$. Thus the isomorphisms $\mathfrak{J}_n$ of $\mathbf{N}_n^{(1)}$ and $\mathbf{N}_n^{(2)}$ for $n = 1, 2, \cdots$ are compatible with each other and they merge to an (algebraical) isomorphism $\mathfrak{J}$ of $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$. In $\mathbf{N}_n^{(1)}$, our $\mathfrak{J}$ agrees with $\mathfrak{J}_n$. Hence $\mathfrak{J}$ maps $\mathbf{N}_n^{(1)}$ on $\mathbf{N}_n^{(2)}$.

In $\mathbf{N}_n^{(h)}$, $Tr_{\mathbf{N}_n^{(h)}}(A)$ is purely algebraical (cf. §1.2). Clearly $Tr_{\mathbf{M}^{(h)}}(A)$ will do as $Tr_{\mathbf{N}_n^{(h)}}(A)$ for the $A \in \mathbf{N}_n^{(h)}$ in the sense of [6], p. 219, Property IV. And as the trace is uniquely determined by that property, so $Tr_{\mathbf{M}^{(h)}}(A) = Tr_{\mathbf{N}_n^{(h)}}(A)$ for $A \in \mathbf{N}_n^{(h)}$. Hence $[[A]] = (Tr_{\mathbf{M}^{(h)}}(A^*A))^{\frac{1}{2}} = (Tr_{\mathbf{N}_n^{(h)}}(A^*A))^{\frac{1}{2}}$ for $A \in \mathbf{N}_n^{(h)}$ and therefore $[[A]]$ and with it $[[A - B]]$ is purely algebraical in $\mathbf{N}_n^{(h)}$.

Thus $\mathfrak{J}$ leaves $[[A - B]]$ invariant in $\mathbf{N}_n^{(h)}$ for all $n = 1, 2, \cdots$. Hence the same is true in all $\mathbf{S}^{(h)}$.

Summing up: $\mathfrak{J}$ is an isomorphism of $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ with respect to

$$1, \alpha A, A^*, A + B, AB, \tag{4.1.γ}$$

and

$$[[A - B]]. \tag{4.1.δ}$$

By (4.1.δ) $\mathfrak{J}$ is isometric. Now extend $\mathfrak{J}$ as far as possible in $\mathbf{B}$ by metric convergence, thus obtaining a new isometric (and hence one-to-one) correspondence $\mathfrak{F}$. By (ii) in Lemma 4.1.4, $\mathfrak{F}$ maps all of $\mathbf{M}^{(1)}$ on $\mathbf{M}^{(2)}$.

The continuity of the operations (4.1.γ) (cf. the proof of (ii) in Lemma 4.1.4) guarantees that $\mathfrak{F}$ too leaves (4.1.γ) invariant. Hence $\mathfrak{F}$ is an (algebraical) isomorphism of $\mathbf{M}_1$ and $\mathbf{M}_2$. This completes the proof.

§**4.2** As a preliminary before introducing other definitions of approximate finiteness, we develop further the ideas which appear in §1.5.

LEMMA 4.2.1. *For any* $\epsilon > 0$ *there exists an* $w_3 = w_3(\epsilon) > 0$ *with the following property*:

Consider a ring $\mathbf{N} \subseteq \mathbf{M}$ *and a projection* $E \epsilon \mathbf{M}$. *If there exists an* $A \epsilon \mathbf{N}$ *with* $[[A - E]] < w_3$ *then there exists also a projection* $F \epsilon \mathbf{N}$ *with* $[[F - E]] < \epsilon$.

PROOF. Consider first the hypothetical A. We have $[[A^* - E]] = [[(A - E)^*]] = [[A - E]] < w_3$. Since $\frac{1}{2}(A + A^*) - E = \frac{1}{2}((A - E) + (A^* - E))$ these inequalities imply $[[\frac{1}{2}(A + A^*) - E]] < w_3$ by the triangle inequality. Thus we may replace A by $\frac{1}{2}(A + A^*)$ i.e. we may assume that A is Hermitian.

Now form the following functions

$$\psi_1(\lambda) \begin{cases} = 1 & \text{for } \quad |\lambda| \geqq 1 \\ = 2|\lambda| - 1 & \text{for } \quad \frac{1}{2} \leqq |\lambda| < 1 \\ = 0 & \text{for } \quad |\lambda| \leqq \frac{1}{2}, \end{cases}$$

$$\psi_2(\lambda) \begin{cases} = 1 & \text{for } \quad |\lambda| \geqq \frac{1}{2} \\ = 2|\lambda| & \text{for } \quad |\lambda| < \frac{1}{2}. \end{cases}$$

Clearly Lemma 1.5.3 applies to both of them. We form the $w_2(\epsilon)$ of that lemma for both functions and denote their minimum by $w_2'(\epsilon)$. We put

$$w_3(\epsilon) = w_2'\left(\frac{\epsilon}{2}\right). \tag{4.2.α}$$

Form the further function

$$\varphi(\lambda) \begin{cases} = 1 & \text{for } \quad |\lambda| \geqq \frac{1}{2} \\ = 0 & \text{for } \quad |\lambda| < \frac{1}{2}.\text{[31]} \end{cases}$$

The following facts are clear:

$\varphi(\lambda) = \overline{\varphi(\lambda)} = \varphi(\lambda)^2$. Hence $F = \varphi(A)$ is a projection which belongs to $\mathbf{M}$ along with A. $\psi_1(\lambda) = \psi_2(\lambda) = \varphi(\lambda) = \lambda$ for $\lambda = 0, 1$, the entire spectrum of E. Hence

$$\psi_1(E) = \psi_2(E) = \varphi(E) = E. \tag{4.2.β}$$

Finally $(1 - \varphi(\lambda))\psi_1(\lambda) + \varphi(\lambda)\psi_2(\lambda) = \varphi(\lambda)$. Hence

$$(1 - F)\psi_1(A) + F\psi_2(A) = F. \tag{4.2.γ}$$

Now $[[A - E]] < w_3(\epsilon)$ implies by Lemma 1.5.3 and (4.2.α) that $[[\psi_1(A) - \psi_1(E)]] < \epsilon/2$ and $[[\psi_2(A) - \psi_2(E)]] < \epsilon/2$. Hence (4.2.$\beta$) now yields that $[[\psi_1(A) - E]] < \epsilon/2$ and $[[\psi_2(A) - E]] < \epsilon/2$. Since $|||\, F \,|||$ and $|||\, 1 - F \,|||$ are $\leqq 1$[32] we can therefore conclude that

$$[[(1 - F)(\psi_1(A) - E) + F(\psi_2(A) - E)]] < \epsilon.$$

[31] Observe that $\psi_1(\lambda)$, $\psi_2(\lambda)$ are continuous, while $\varphi(\lambda)$ is not.

[32] Actually we have an equality in the case of each of these operators except when it is zero.

This may be rewritten

$$[[(1 - F)\psi_1(A) + F\psi_2(A) - E]] < \epsilon.$$

With (4.2.γ) this gives $[[F - E]] < \epsilon$ as desired.

LEMMA 4.2.2. *For any $\epsilon > 0$, there exists an $w_4 = w_4(\epsilon) > 0$ with the following property.*

Consider two projections $E, F \in \mathbf{M}$ with the closed linear sets $\mathfrak{M}, \mathfrak{N}$. If $[[E - F]] < w_4$, then there exists a partially isometric $W' \in \mathbf{M}$ with an initial set $\mathfrak{M}' \subseteq \mathfrak{M}$ a final set $\mathfrak{N}' \subseteq \mathfrak{N}$ and with $[[W' - E]] < \epsilon$.

PROOF. Form the canonical decomposition of $A = FE$ in the sense of [5], p. 143, Def. 4.4.1:

$$FE = A = W'B. \qquad EF = A^* = BW'^*. \tag{4.2.δ}$$

Here B is Hermitian and definite, W' is partially isometric with initial set $\mathfrak{M}' =$ [Range A^*] $=$ [Range EF] $\subseteq$ [Range E] $= \mathfrak{M}$ and final set $\mathfrak{N}' =$ [Range A] $=$ [Range FE] $\subseteq$ [Range F] $= \mathfrak{N}$.

The range of W'^* is $\mathfrak{M}' \subset \mathfrak{M} =$ [Range E]. Thus $EW'^* = W'^*$ and taking adjoints, we obtain

$$W'E = W'. \tag{4.2.ϵ}$$

For any canonical decomposition $B^2 = A^*A$. In our case $A^*A = EF \cdot FE = EFE$ so

$$B^2 = EFE \tag{4.2.ζ}$$

Observe also that due to the character of these operators

$$||| E ||| \leqq 1, \qquad ||| F ||| \leqq 1, \qquad ||| W' ||| \leqq 1.^{32} \tag{4.2.η}$$

Thus (4.2.ζ) implies $||| B^2 ||| \leqq 1$ and so

$$||| B ||| \leqq 1. \tag{4.2.θ}$$

Now form the function

$$\psi(\lambda) \begin{cases} = 1 & \text{for } |\lambda| \geqq 1 \\ = \sqrt{|\lambda|} & \text{for } |\lambda| \leqq 1. \end{cases}$$

Clearly Lemma 1.5.3 applies to it. We form the $w_2(\epsilon)$ of that lemma for this function and put

$$w_4(\epsilon) = \text{Min}\left(w_2\left(\frac{\epsilon}{2}\right), \frac{\epsilon}{2}\right). \tag{4.2.ι}$$

It is clear that $\psi(\lambda^2) = \lambda$ for $0 \leqq \lambda \leqq 1$ which includes the entire spectrum of B, since B is definite and (4.2.θ) holds. Thus $\psi(B^2) = B$, or using (4.2.ζ) we obtain

$$\psi(EFE) = B. \tag{4.2.κ}$$

On the other hand, $\psi(\lambda) = \lambda$ for $\lambda = 0, 1$ the spectrum of E and hence

$$\psi(E) = E. \tag{4.2.λ}$$

Now our assumption, $[[E - F]] < w_4$ implies by (4.2.η) that $[[(E - F)E]] < w$ and $[[E(E - F)E]] < w_4$ or

$$[[E - FE]] < w_4, \qquad [[E - EFE]] < w_4 \tag{4.2.μ}$$

(4.2.μ), Lemma 1.5.3 and (4.2.ι) together imply

$$[[\psi(E) - \psi(EFE)]] < \frac{\epsilon}{2}.$$

This and (4.2.κ) and (4.2.λ) yield $[[E - B]] < \epsilon/2$. This and (4.2.η) imply $[[W'(E - B)]] < \epsilon/2$. Hence (4.2.$\epsilon$) and (4.2.$\delta$) give

$$[[W' - FE]] < \frac{\epsilon}{2}. \tag{4.2.ν}$$

(4.2.μ) and (4.2.ι) imply $[[E - FE]] < \epsilon/2$. Hence (4.2.ν) yields $[[W' - E]] < \epsilon$ as desired.

LEMMA 4.2.3. *If $U \in \mathbf{M}$ is unitary and $W \in \mathbf{M}$ is partially isometric, and if U agrees with W on its initial set, then*

$$[[U - W]] \leqq (2\,[[W - 1]])^{\frac{1}{2}}$$

PROOF. The initial set of W is the range or final set of W^*. Therefore $UW^* = WW^*$. This implies for the adjoints $WU^* = WW^*$. Hence

$$\begin{aligned}(U - W)(U - W)^* &= UU^* - UW^* - WU^* + WW^* \\ &= 1 - WW^* - WW^* + WW^* = 1 - WW^*.\end{aligned}$$

Therefore $[[U - W]]^2 = Tr_{\mathbf{M}}((U - W)(U - W)^*) = Tr_{\mathbf{M}}(1 - WW^*)$ or

$$[[U - W]] = (Tr_{\mathbf{M}}(1 - WW^*))^{\frac{1}{2}} \tag{4.2.o}$$

We now make use of Schwartz's inequality in the sense of [6], p. 241, Theorem VIII, remembering that $|||\, W\, ||| \leqq 1$ and so $[[W]] \leqq 1$. Thus

$$Tr_{\mathbf{M}}(1 - WW^*) = Tr_{\mathbf{M}}(-W(W - 1)^* - (W - 1)1^*) \leqq 2[[W - 1]].$$

This and (4.2.o) yield the desired inequality of the lemma.

LEMMA 4.2.4. *For any $\epsilon > 0$, there exists an $w_5 = w_5(\epsilon) > 0$ with the following property:*

Consider two projections, $E, F \in \mathbf{M}$ with $D_{\mathbf{M}}(E) = D_{\mathbf{M}}(F)$. If $[[E - F]] < w_5$, then there exists a unitary $U \in \mathbf{M}$ with $F = UEU^{-1}$ and with $[[U - 1]] < \epsilon$.

PROOF. Put $w_5(\epsilon) = w_4(\epsilon')$ where $\epsilon' = \epsilon'(\epsilon) > 0$ will be chosen later.

Apply Lemma 4.2.2 to E, F and their closed linear sets $\mathfrak{M}$, $\mathfrak{N}$. It gives a partially isometric $W' \in \mathbf{M}$ with an initial set $\mathfrak{M}' \subseteq \mathfrak{M}$ and a final set $\mathfrak{N}' \subseteq \mathfrak{N}$ and with

$$[[W' - E]] < \epsilon'. \tag{4.2.ξ}$$

In view of the relation of W' to $\mathfrak{M}'$ and $\mathfrak{N}'$, $D_{\mathbf{M}}(\mathfrak{M}') = D_{\mathbf{N}}(\mathfrak{N}')$. By hypothesis, $D_{\mathbf{M}}(\mathfrak{M}) = D_{\mathbf{N}}(\mathfrak{N})$. Hence $D_{\mathbf{M}}(\mathfrak{M} \ominus \mathfrak{M}') = D_{\mathbf{M}}(\mathfrak{N} \ominus \mathfrak{N}')$. Choose a partially isometric $V' \in \mathbf{M}$ with initial set $\mathfrak{M} \ominus \mathfrak{M}'$ and final set $\mathfrak{N} \ominus \mathfrak{N}'$. Then $U' = W' + V' \in \mathbf{M}$ is also partially isometric, with initial set $\mathfrak{M}$, final set $\mathfrak{N}$, and it agrees with W' on the initial set $\mathfrak{M}'$ of W.

Apply all this to $1 - E$, $1 - F$ and $\mathfrak{M}^{\perp}(= \mathfrak{H} \ominus \mathfrak{M})$, $\mathfrak{N}^{\perp}$ in place of E, F and $\mathfrak{M}$, $\mathfrak{N}$. (Clearly $[[(1 - E) - (1 - F)]] = [[E - F]]$). Thus we obtain a partially isometric $W'' \in \mathbf{M}$ with an initial set $\mathfrak{M}'' \subseteq \mathfrak{M}^{\perp}$ a final set $\mathfrak{N}'' \subseteq \mathfrak{N}^{\perp}$ and with

$$(4.2.\pi) \qquad [[W'' - (1 - E)]] < \epsilon'.$$

Furthermore, there exists a partially isometric $U'' \in \mathbf{M}$ with the initial set $\mathfrak{M}^{\perp}$ final set $\mathfrak{N}^{\perp}$ and agreeing with W'' on the initial set $\mathfrak{M}''$ of W''.

Now put

$$W = W' + W'', \qquad U = U' + U''.$$

Then $W, U \in \mathbf{M}$ are partially isometric, with initial sets $[\mathfrak{M}', \mathfrak{M}'']$, $[\mathfrak{M}, \mathfrak{M}^{\perp}] = \mathfrak{H}$ respectively, and with the respective final sets $[\mathfrak{N}', \mathfrak{N}'']$ and $[\mathfrak{N}, \mathfrak{N}^{\perp}] = \mathfrak{H}$. So U is unitary.

U agrees with U' on the initial set $\mathfrak{M}$ of U'. Hence it maps $\mathfrak{M}$ on $\mathfrak{N}$. Therefore

$$(4.2.\rho) \qquad F = UEU^{-1}.$$

U agrees with W on the initial set $[\mathfrak{M}', \mathfrak{M}'']$ of W. Therefore by Lemma 4.2.3,

$$(4.2.\sigma) \qquad [[U - W]] = (2[[W - 1]])^{\frac{1}{2}}$$

(4.2.ξ) and (4.2.π) imply

$$(4.2.\tau) \qquad [[W - 1]] < 2\epsilon'$$

(4.2.σ) and (4.2.τ) give together $[[U - W]] < (4\epsilon')^{\frac{1}{2}}$. This and (4.2.$\tau$) yield

$$(4.2.\upsilon) \qquad [[U - 1]] < (4\epsilon')^{\frac{1}{2}} + 2\epsilon'$$

Hence if we choose ϵ' so that $(4\epsilon')^{\frac{1}{2}} + 2\epsilon' \leqq \epsilon$, (4.2.$\upsilon$) yields

$$(4.2.\varphi) \qquad [[U - 1]] < \epsilon$$

In view, therefore, of (4.2.ρ) and (4.2.φ), the w_5 which we have specified satisfies the condition of our lemma.

§4.3 We shall recast the definition of approximate finiteness in a series of steps, i.e. Defs. 4.3.1, 4.5.2 and 4.6.1, below:

DEFINITION 4.3.1. $\mathbf{M}$ *is approximately finite* (A) *if this is true:* *Given any* $A_1, \cdots, A_m \in \mathbf{M}$ *and any* $\epsilon > 0$, *there exists an* $n = n_1(A_1, \cdots,$

$A_m, \epsilon)$ *such that for every* $q \geqq n$ *there exists a factor* $\mathbf{N} = \mathbf{N}_1(q, A_1, \cdots, A_m, \epsilon)$ *with the following properties*:

(i) $\mathbf{N}$ *is in case* (I_q)

(ii) $\mathbf{N} \subseteq \mathbf{M}$

(iii) *There exist* $B_1, \cdots, B_m \in \mathbf{N}$ *with*

$$[[B_h - A_h]] < \epsilon \quad \text{for} \quad h = 1, \cdots, m.$$

In the present section we are interested in certain consequences of this definition, in particular what other properties of the approximating factor $\mathbf{N}$ may be assumed. It will be assumed therefore in the five lemmas of this section that $\mathbf{M}$ is approximately finite (A).

LEMMA 4.3.1. *Given any* $A_1, \cdots, A_m \in \mathbf{M}$ *any projection* $E \in \mathbf{M}$ *with* $D_{\mathbf{M}}(E) = 1/p$ *for* $p = 1, 2, \cdots$ *and any* $\epsilon > 0$, *there exists an* $n_2 = n_2(A_1, \cdots, A_m, E, \epsilon)$ *such that for every* $q \geqq n_2$ *which is divisible by* p *there exists a factor* $\mathbf{N} = \mathbf{N}_2(q, A_1, \cdots, A_m, E, \epsilon)$ *with the following properties*:

(i) $\mathbf{N}$ *is in case* (I_q)

(ii) $\mathbf{N} \subseteq \mathbf{M}$.

(iii) *There exist* $B_1, \cdots, B_m \in \mathbf{N}$ *with*

$$[[B_h - A_h]] < \epsilon \quad \textit{for} \quad h = 1, \cdots, m.$$

(iv) *There exists a projection* $F \in \mathbf{N}$ *with* $D_{\mathbf{M}}(F) = D_{\mathbf{M}}(E) = 1/p$, *and* $[[F - E]] < \epsilon$.[33]

PROOF. Apply Def. 4.3.1 with $A_1, \cdots, A_m, E$ in place of its $A_1, \cdots, A_m$ and with ϵ' in place of ϵ where $\epsilon' = \epsilon'(\epsilon)$ will be chosen later.

Put $n_2(A_1, \cdots, A_m, E, \epsilon) = n_1(A_1, \cdots, A_m, E, \epsilon')$ and $\mathbf{N}_2(q, A_1, \cdots, A_m, E, \epsilon) = \mathbf{N}_1(q, A_1, \cdots, A_m, E, \epsilon)$. Now consider a $q \geqq n_2$ which is divisible by p and its $\mathbf{N} = \mathbf{N}_2$.

In view of the above definition of $\mathbf{N}_2$, (i) and (ii) of the present lemma follow from (i) and (ii) of Def. 4.3.1. Furthermore, (iii) of Def. 4.3.1 states that

$$[[B_h - A_h]] < \epsilon' \quad \text{for} \quad h = 1, \cdots, m, \tag{4.3.α}$$

and

$$[[B_{m+1} - E]] < \epsilon'. \tag{4.3.β}$$

Put

$$\epsilon' = \text{Min.}\ (w_3(\epsilon''), \epsilon) \tag{4.3.γ}$$

where $\epsilon'' = \epsilon''(\epsilon)$ will be chosen later. Now (4.3.α) and (4.3.γ) imply (iii) in the lemma, so we must only derive (iv) from (4.3.β).

---

[33] Observe that this is a strengthened form of Def. 4.3.1.

By Lemma 4.2.1, (4.3.β) and (4.3.γ) imply the existence of a projection $F_1 \epsilon \mathbf{N}$ with

$$[[F_1 - E]] < \epsilon''. \tag{4.3.δ}$$

This implies

$$| D_{\mathbf{M}}(F_1) - D_{\mathbf{M}}(E) | < \epsilon''.^{34} \tag{4.3.ϵ}$$

As $\mathbf{N}$ is in case (I_q) and q is divisible by p, there exists a projection $F \epsilon \mathbf{N}$ with

$$D_{\mathbf{M}}(F) = D_{\mathbf{M}}(E) = 1/p, \tag{4.3.ζ}$$

and we can even choose $F_1 \gtreqless F$. This latter implies $[[F_1 - F]] = (| D_{\mathbf{M}}(F_1) - D_{\mathbf{M}}(F) |)^{\frac{1}{2}}$.[35] This equation, (4.3.ϵ) and (4.3.ζ) together imply $[[F_1 - F]] < \epsilon''^{\frac{1}{2}}$. Hence (4.3.$\delta$) now implies

$$[[F - E]] < \epsilon''^{\frac{1}{2}} + \epsilon''. \tag{4.3.η}$$

Thus if we choose $\epsilon'' = \epsilon''(\epsilon) > 0$, so that $(\epsilon'')^{\frac{1}{2}} + \epsilon'' \leqq \epsilon$ we can conclude from (4.3.η) that $[[F - E]] < \epsilon$. In view of (4.3.ζ) and $F \epsilon \mathbf{N}$, our last inequality shows that $\mathbf{N}$ has property (iv) as desired.

LEMMA 4.3.2. *Given any $A_1, \cdots, A_m \epsilon \mathbf{M}$ any projection $E \epsilon \mathbf{M}$ with $D_{\mathbf{M}}(E) = 1/p$ for $p = 1, 2, \cdots$ and any $\epsilon > 0$. Then there exists an $n_3 = n_3(A_1, \cdots, A_m, E, \epsilon)$ such that for every $q \geqq n_3$ which is divisible by p there exists a factor $\mathbf{N} = \mathbf{N}_3(q, A_1, \cdots, A_m, E, \epsilon)$ with the properties* (i), (ii), (iii) *of Lemma* 4.3.1, *and in addition*

(iv′) $E \epsilon \mathbf{N}$.[33]

PROOF. Our hypotheses are identical with those of Lemma 4.3.1, and we apply the latter to $A_1, \cdots, A_m, E$ but with ϵ' in place of ϵ where

$$\epsilon' = \text{Min.}\ (w_5(\epsilon''), \epsilon'') \tag{4.3.θ}$$

and $\epsilon'' = \epsilon''(\epsilon) > 0$ will be chosen later.

Denote the n_2, $\mathbf{N}_2$ and $B_1, \cdots, B_m, F$ which we obtain in this way by n_2',

---

[34] For any two projections $E, F \epsilon \mathbf{M}$

$$| D_{\mathbf{M}}(F) - D_{\mathbf{M}}(E)| \leqq [[F - E]].$$

Proof. We have

$$| D_{\mathbf{M}}(F) - D_{\mathbf{M}}(E)| = | Tr_{\mathbf{M}}(F) - Tr_{\mathbf{M}}(E)| = | Tr_{\mathbf{M}}(F - E)|$$

This is $\leqq [[1]]\cdot[[F - E]]$ (cf. [6], pp. 241, Theorem VIII); hence $\leqq [[F - E]]$ as desired.

[35] For any two projections $E, F \epsilon \mathbf{M}$ with $E \gtreqless F$

$$[[F - E]] = (| D_{\mathbf{M}}(F) - D_{\mathbf{M}}(E)|)^{\frac{1}{2}}$$

Proof. By symmetry we may assume $E \leqq F$. Then $F - E$ is a projection, $D_{\mathbf{M}}(F) - D_{\mathbf{M}}(E) = D_{\mathbf{M}}(F - E)$. Put $G = F - E$. We must prove $[[G]] = (D_{\mathbf{M}}(G))^{\frac{1}{2}}$. Now

$$[[G]]^2 = Tr_{\mathbf{M}}(G^*G) = Tr_{\mathbf{M}}(G) = D_{\mathbf{M}}(G).$$

$\mathbf{N}_2^+$ and $B_1', \cdots, B_m', F'$. We put $n_3 = n_2'$ and $\mathbf{N}^+ = \mathbf{N}_2^+$. In view of (4.3.$\theta$), Lemma 4.3.1 (iii) gives

$$(4.3.\iota) \qquad [[B_h' - A_h]] < \epsilon'' \quad \text{for} \quad h = 1, \cdots, m,$$

while (iv) of that Lemma yields

$$(4.3.\kappa) \qquad [[F' - E]] < w_5(\epsilon'')$$

and

$$(4.3.\lambda) \qquad D_{\mathbf{M}}(F') = D_{\mathbf{M}}(E).$$

(4.3.κ) and (4.3.λ) allow us to apply Lemma 4.2.4 with F', E in place of its E, F. Consequently there exists a unitary $U \in \mathbf{M}$ with

$$(4.3.\mu) \qquad [[U - 1]] < \epsilon''$$

and

$$(4.3.\nu) \qquad E = UF'U^{-1}.$$

(4.3.μ) implies $[[A_h - U^{-1}A_hU]] = [[A_h - A_hU + A_hU - U^{-1}A_hU]] \leqq ||| A_h ||| \, [[1 - U]] + [[1 - U^{-1}]] \, ||| A_hU ||| \leqq 2 \, ||| A_h ||| \cdot \epsilon''$. Hence by (4.3.$\iota$) $[[B_h' - U^{-1}A_hU]] < (2 \, ||| A_h ||| + 1)\epsilon''$ and so

$$(4.3.o) \qquad [[UB_h'U^{-1} - A_h]] < (2 \, ||| A_h ||| + 1)\epsilon''.$$

Now we choose $\epsilon'' = \epsilon''(\epsilon) > 0$ as

$$(4.3.\xi) \qquad \epsilon'' = \epsilon/(2 \operatorname{Max}_{h=1,\cdots,m} ||| A_h ||| + 1).$$

Put $\mathbf{N} = \mathbf{N}_3 = U\mathbf{N}^+U^{-1}$. Then $\mathbf{N}$ is a factor in the case (I_q) along with $\mathbf{N}^+$. Thus (i) holds, (ii) holds since $\mathbf{N}^+ \subseteq \mathbf{M}$. $U \in \mathbf{M}$ implies $\mathbf{N} \subseteq \mathbf{M}$. $B_h' \in \mathbf{N}^+$ gives $B_h = UB_hU^{-1} \in \mathbf{N}$, while (4.3.$o$) and (4.3.$\xi$) imply $[[B_h - A_h]] < \epsilon$, so (iii) holds. And finally $F' \in \mathbf{N}^+$ and (4.3.ν) imply $E = UF'U^{-1} \in \mathbf{N}$ so that (iv$'$) holds.

LEMMA 4.3.3. *If we add to the assumptions of Lemma 4.3.2 that*

$$(\alpha) \qquad EA_h = A_hE = A_h$$

then we can add to its assertions that

$$(\beta) \qquad EB_h = B_hE = B_h.$$ [36]

PROOF. Apply Lemma 4.3.2, but write B_h' for its B_h. Now put $B_h = EB_h'E$. Then (β) is satisfied and $B_h', E \in \mathbf{N}$ give $B_h \in \mathbf{N}$. By (α), $EA_hE = A_h$. By (iii) of Lemma 4.3.2, $[[B_h' - A_h]] < \epsilon$. Hence $[[E(B_h' - A_h)E]] < \epsilon$ or $[[B_h - A_h]] < \epsilon$. Thus (iii) holds for B_h as well as for B_h'.

The other assertions ((i), (ii), (iv$'$)) are not affected by our replacing B_h' by B_h. Thus the proof is completed.

[36] With the same n_3, N_3 as in that lemma.

LEMMA 4.3.4. *If we add to the assumption of Lemma* 4.3.3 *that* $E \in \mathbf{N}^0$ *where* $\mathbf{N}^0$ *is a factor in case* (I_p), *then we can add to its assertions*

(v) $\mathbf{N}^0 \subseteq \mathbf{N}$.[33, 36]

PROOF. Apply Lemma 4.3.3 but write $\mathbf{N}^+$ for its $\mathbf{N}$.

The considerations at the beginning of the proof of (i) in Lemma 4.1.2 show again that $D_{\mathbf{M}}(G) = D_{\mathbf{N}^0}(G)$ for all projections $G \in \mathbf{N}^0 (\subseteq \mathbf{M})$.

Hence in particular $D_{\mathbf{N}_0}(E) = D_{\mathbf{M}}(E) = 1/p$. Then the argument used in proof of Lemma 2.6.2 establishes the existence of a system of p-order matrix units $W^0_{u,v}$, $u, v = 1, \cdots, p$ in $\mathbf{N}^0$ with $\sum_{u=1}^p W^0_{u,u} = 1$ and $W^0_{1,1} = E$. Thus in terms of (β) of Lemma 4.1.1,

$$(4.3.\pi) \quad \text{There exists a matrix base } W^0_{u,v}, \quad u, v = 1, \cdots, p \text{ of } \mathbf{N}^0 \text{ with } W^0_{1,1} = E.$$

On the other hand, the argument of Lemma 2.6.3 applied to $\mathbf{N}^+$ establishes the existence of a system of p-order matrix units $W^+_{u,v}$ $u, v = 1, \cdots, p$ in $\mathbf{N}^+$ with $\sum_{u=1}^p W^+_{u,u} = 1$ and $W^+_{1,1} = E$. (Observe that this is not a matrix base of $\mathbf{N}^+$ since $\mathbf{N}^+$ is in case (I_q) and not (I_p)!).

Now put $U = \sum_{u=1}^p W^0_{u,1} W^+_{1,u}$. As $W^0_{u,1} \in \mathbf{N}^0 \subseteq \mathbf{M}$, $W^+_{1,u} \in \mathbf{N} \subseteq \mathbf{M}$ so $U \in \mathbf{M}$. One verifies by direct computation that $U^*U = UU^* = 1$, i.e. U is unitary. Also

$$(4.3.\rho) \qquad UW^+_{u,v} = W^0_{u,v}U \quad \text{or} \quad W^0_{u,v} = UW^+_{u,v}U^{-1}$$

and, since $W^0_{1,1} = W^+_{1,1} = E$

$$(4.3.\sigma) \qquad EU = UE = E.$$

Apply the mapping

$$(4.3.\tau) \qquad X \to UXU^{-1}$$

to $\mathbf{N}^+$. This carries $\mathbf{N}^+$ into $\mathbf{N} = U\mathbf{N}^+U^{-1}$. Now we see: $\mathbf{N}$ is a factor in case (I_q) along with $\mathbf{N}^+$; thus (i) holds. $\mathbf{N}^+ \subseteq \mathbf{M}$, $U \in \mathbf{M}$ imply $\mathbf{N} \subseteq \mathbf{M}$; so (ii) holds. By (β) in Lemma 4.3.3 and $(4.3.\sigma)$, the mapping $(4.3.\tau)$ leaves all B_h fixed; hence they belong to $\mathbf{N}$ as well as to $\mathbf{N}^+$. Now (iii) and (β) in Lemma 4.3.2 are unaffected and hence hold in the new case. By $(4.3.\rho)$ the mapping $(4.3.\tau)$ carries $W^+_{u,v}$ into $W^0_{u,v}$. Hence $W^+_{u,v} \in \mathbf{N}^+$ gives $W^0_{u,v} \in \mathbf{N}$. Hence by $(4.3.\pi)$ and (ii) in Lemma 4.1.1, every $A \in \mathbf{N}^0$ has the form $A = \sum_{u=1}^p \sum_{v=1}^p a_{v,u} W^0_{u,v}$ so $A \in \mathbf{N}$. Thus $\mathbf{N}^0 \subseteq \mathbf{N}$ which is the desired relation (v).

Thus the proof is completed.

LEMMA 4.3.5. *Assume that* $\mathbf{M}$ *is approximately finite* (A). *Given any* $A_1, \cdots, A_m \in \mathbf{M}$ *any* $p = 1, 2, \cdots$ *and any* $\epsilon > 0$, *then there exists an* $n_4 = n_4(A_1, \cdots, A_m, p, \epsilon)$ *such that for every* $q \geq n_4$ *which is divisible by* p *and every*

factor $\mathbf{N}^0 \subseteq \mathbf{M}$ *in case* (I_p), *there exists a factor* $\mathbf{N} = \mathbf{N}_4(q, A_1, \cdots, A_m, \mathbf{N}^0, p, \epsilon)$ *with the following properties*:

(i) $\mathbf{N}$ *is in case* (I_q).

(ii) $\mathbf{N} \subseteq \mathbf{M}$.

(iii) *There exist* $B_1, \cdots, B_m \in \mathbf{N}$ *with* $[[B_h - A_h]] < \epsilon$ *for* $h = 1, \cdots, m$.

(v) $\mathbf{N}^0 \subseteq \mathbf{N}$.

(*Note.* This lemma is quite similar to Lemma 4.3.4. However, (α) has been dropped from the assumptions and (β) from the conclusion. Thus E is no longer mentioned.)

PROOF. Choose a matrix base $W_{u,v}$, $u, v = 1, \cdots, p$ of $\mathbf{N}^0$ and let $E = W_{1,1}$. Put $A_{u,v;h} = W_{1,u}A_hW_{v,1}$.

We have $D_{\mathbf{M}}(E) = D_{\mathbf{M}}(W_{1,1}) = 1/p$ (cf. the argument of (2.7.α)). Clearly $EA_{u,v;h} = A_{u,v;h}E = A_{u,v;h}$.

Now apply Lemma 4.3.4 with the mp^2 operators $A_{u,v;h}$ $h = 1, \cdots, m$, $u, v = 1, \cdots, p$ in place of $A_1, \cdots, A_m$ and ϵ/p^2 in place of ϵ.

Denote the m_3, $\mathbf{N}_3$ which obtain in this way by n_4, $\mathbf{N}_4$. Write $\mathbf{N} = \mathbf{N}_4$. Denote the operators which correspond to the $A_{u,v;h}$ (as the B_h do to the A_h in Lemma 4.3.4) by $B_{u,v;h}$.

Clearly:

$$A_h = \sum_{u=1}^{p} \sum_{v=1}^{p} W_{u,1}A_{u,v;h}W_{1,v}. \tag{4.3.υ}$$

Let

$$B_h = \sum_{u=1}^{p} \sum_{v=1}^{p} W_{u,1}B_{u,v;h}W_{1,v}. \tag{4.3.φ}$$

Now we argue as follows:

(i), (ii) and (v) in Lemma 4.3.4 are unaffected. $B_{u,v;h} \in \mathbf{N}$ and $W_{u,v} \in \mathbf{N}^0 \subseteq \mathbf{N}$ imply $B_h \in \mathbf{N}$. We have $[[B_{u,v;h} - A_{u,v;h}]] < \epsilon/p^2$. Hence $[[W_{u,1}(B_{u,v;h} - A_{u,v;h})W_{1,v}]] < \epsilon/p^2$ and $[[\sum_{u=1}^{p} \sum_{v=1}^{p} W_{u,1}(B_{u,v;h} - A_{u,v;h})W_{1,v}]] < \epsilon$. Thus Thus (4.3.$\upsilon$) and (4.3.$\varphi$) and this inequality yields $[[B_h - A_h]] < \epsilon$. So (iii) holds and the proof is complete.

§**4.4** We now use the results of the preceding two sections to compare the two notions of approximate finiteness which we have introduced.

LEMMA 4.4.1. *Assume that* $\mathbf{M}$ *is approximately finite* (A) *and that the sequence* $[p_1, p_2, \cdots]$ *fulfills* (*i*), (*ii*) *in Lemma* 4.1.3. *Then* $[p_1, p_2, \cdots]$ *possesses a subsequence* $[p_{1'}, p_{2'}, \cdots]$ *such that* $\mathbf{M}$ *is approximately finite of type* $[p_{1'}, p_{2'}, \cdots]$.

PROOF. By [2], pp. 387–388, there exists a sequence $A_1, A_2, \cdots \in \mathbf{M}$ which is everywhere dense in $\mathbf{M}$ in the sense of strong convergence. Put $p_0 = 1$ and $\mathbf{N}_0 = (\alpha 1)$. This is a factor $\subseteq \mathbf{M}$ in case (I_1). We shall choose by induction for every $n = 1, 2, \cdots$, a $p_{n'}$ and a factor $\mathbf{N}_n \subseteq \mathbf{M}$ in the case $(I_{p_{n'}})$. If for any $n = 1, 2, \cdots, p_{(n-1)'}$ and $\mathbf{N}_{n-1}$ have already been chosen, then we choose $p_{n'}$ and $\mathbf{N}_n$ in the following way. Apply Lemma 4.3.5 to $A_1, \cdots, A_n, p_{(n-1)'}$, $1/n$, $\mathbf{N}_{n-1}$ in place of its $A_1, \cdots, A_m, p, \epsilon, \mathbf{N}^0$. Choose $p_{n'}$ from the sequence $[p_1, p_2, \cdots]$ with $p_{n'} > \mathrm{Max}\,(n_4, p_{(n-1)'})$. Thus $p_{n'}$ is divisible by $p_{(n-1)'}$.

Put $q = p_{n'}$ and then put $\mathbf{N}_n = \mathbf{N}$ (from the above application of Lemma 4.3.5).

Thus $[p_{1'}, p_{2'}, \cdots]$ is a subsequence of $[p_1, p_2, \cdots]$. We have $\mathbf{N}_{n-1} \subseteq \mathbf{N}_n$, hence

$$(4.4.\alpha) \qquad \mathbf{N}_1 \subseteq \mathbf{N}_2 \subseteq \cdots .$$

$\mathbf{N}_n$ is a factor $\subseteq \mathbf{M}$ of class $(I_{p_{n'}})$. Clearly

$$(4.4.\beta) \qquad \mathbf{R}(\mathbf{N}_n, n = 1, 2, \cdots) \subseteq \mathbf{M}.$$

Now for $n \geqq h$, there exists a $B_h^{(n)} \in \mathbf{N}_n$ with $[[B_h^{(n)} - A_h]] < 1/n$. So A_h is a metric limit of $\mathbf{R}(\mathbf{N}_n, n = 1, 2, \cdots)$ and hence, by Theorem I, $A_h \in \mathbf{R}(\mathbf{N}_n, n = 1, 2, \cdots)$. Thus $A_1, A_2, \cdots \in \mathbf{R}(\mathbf{N}_n, n = 1, 2, \cdots)$. Since the $A_1, A_2, \cdots$ are everywhere dense in $\mathbf{M}$ in the sense of the strong convergence, this and (4.4.β) yield

$$(4.4.\gamma) \qquad \mathbf{R}(\mathbf{N}_n, n = 1, 2, \cdots) = \mathbf{M}.$$

(4.4.α) and (4.4.γ) and our observations establish that $\mathbf{M}$ is approximately finite of type $[p_{1'}, p_{2'}, \cdots]$ as stated.

We need one more auxiliary lemma. In view of a later application, we shall prove it in a stronger form than that immediately needed.

LEMMA 4.4.2. *Let* $\mathbf{N}$, $\mathbf{O}$ *be two factors in the case* (I_p) *and* (I_q) *or* (II_1) *respectively. Let p be a divisor of r and (if q exists) r a divisor of q. Suppose* $\mathbf{N} \subseteq \mathbf{O}$. *Then there exists a factor* $\mathbf{R}$ *in case* (I_r) *such that* $\mathbf{N} \subseteq \mathbf{R} \subseteq \mathbf{O}$.

PROOF. By Lemma 2.6.2 or Lemma 2.6.3, since $\mathbf{O}$ is in case (II_1) or in the case (I_q), q divisible by r, there exists a factor $\mathbf{R} \subseteq \mathbf{O}$ in the case (I_r) with a matrix base $W^+_{o,w}$, $o, w = 1, \cdots, r$. Choose also a matrix base $W^0_{u,v}$, $u,v = 1, \cdots, p$ of N^+. Thus $W^+_{o,w}$ is a system of r-order matrix units with $\sum_{o=1}^{r} W^+_{o,o} = 1$ and $W^0_{u,v}$ is a system of p-order matrix units with $\sum_{u=1}^{p} W^0_{u,u} = 1$.

Put

$$(4.4.\delta) \qquad W^{++}_{u,v} = \sum_{i=0}^{r/p} W^+_{(u-1)r/p+i,(v-1)r/p+i} .$$

Then one verifies by direct calculation that the $W^{++}_{u,v}$, $u, v = 1, \cdots, p$ form a system of p-order units with $\sum_{u=1}^{p} W^{++}_{u,u} = 1$. One can readily prove that $D_{\mathbf{O}}(W^{++}_{1,1}) = 1/p = D_{\mathbf{O}}(W^0_{1,1})$ in the standard normalization. Thus there exists a partially isometric $V \in \mathbf{O}$ such that $V^*V = W^{++}_{1,1}$, $VV^* = W^0_{1,1}$. Now put

$$U = \sum_{u=1}^{p} W^0_{u,1} V W^{++}_{1,u} .$$

Clearly $U \in \mathbf{O}$. One verifies by direct computation that $U^*U = UU^* = 1$ and hence that U is unitary. Also that

$$(4.4.\epsilon) \qquad UW^{++}_{u,v} = W^0_{u,v}U \quad \text{or} \quad W^0_{u,v} = UW^{++}_{u,v}U^{-1}.$$

Hence the replacement of $\mathbf{R}$ by $U\mathbf{R}U^{-1}$ and of the system $W^+_{o,w}$ by the system $UW^+_{o,w}U^{-1}$ leaves all their properties unaffected, but it carries $W^{++}_{u,v}$ into $W^0_{u,v} = UW^{++}_{u,v}U^{-1}$. Consequently, we may assume that we had $W^{++}_{u,v} = W^0_{u,v}$ i.e.

$$(4.4.\zeta) \qquad W^0_{u,v} = \sum_{i=1}^{r/p} W^+_{(u-1)r/p+i,(v-1)r/p+i} .$$

to begin with.

Thus all $W^0_{u,v} \in \mathbf{R}$, hence $\mathbf{N} \subseteq \mathbf{R}$. We know already that $\mathbf{R}$ is a factor in the case (I_r) and that $\mathbf{R} \subseteq \mathbf{O}$.

Thus the proof is completed.

We can now conclude our deductions from Def. 4.3.1.

LEMMA 4.4.3. *Under the same assumptions as Lemma 4.4.1, $\mathbf{M}$ is approximately finite of type $[p_1, p_2, \cdots]$.*

PROOF. Apply Lemma 4.4.1. Denote its $\mathbf{N}_1, \mathbf{N}_2, \cdots$ by $\mathbf{N}_{1'}, \mathbf{N}_{2'}, \cdots$. Write $0' = 0$, $\mathbf{N}_{0'} = \mathbf{N}_0 = (\alpha \cdot 1)$.

Thus $\mathbf{N}_n$ is now defined for $n = 0', 1', 2', \cdots$ only. Consider a fixed $m' = 1', 2', \cdots$. We have $\mathbf{N}_{(m-1)'} \subseteq \mathbf{N}_{m'}$ and that $\mathbf{N}_{(m-1)'}, \mathbf{N}_{m'}$ are factors in the cases $(\mathrm{I}_{p_{(m-1)'}})$, $(\mathrm{I}_{p_{m'}})$ respectively. Each $p_{(m-1)'}, p_{(m-1)'+1}, \cdots, p_{m'-1}, p_{m'}$ is a divisor of its successor. Hence we obtain by repeated applications of Lemma 4.4.2 a succession of factors $\mathbf{N}_{(m-1)'+1}, \mathbf{N}_{(m-1)'+2}, \cdots, \mathbf{N}_{m'-1}$, with these properties:

$$\mathbf{N}_{(m-1)'} \subseteq \mathbf{N}_{(m-1)'+1} \subseteq \cdots \subseteq \mathbf{N}_{m'-1} \subseteq \mathbf{N}_{m'}$$

and $\mathbf{N}_n$ $(n = (m-1)' + 1, \cdots, m' - 1)$ is in the case (I_{p_n}).

So we have defined $\mathbf{N}_n$ for all $n = 1, 2, \cdots$. We have

$$\mathbf{N}_1 \subseteq \mathbf{N}_2 \subseteq \cdots, \tag{4.4.η}$$

and $\mathbf{N}_n$ is a factor in the case (I_{p_n}). Finally, by (4.4.η) and Lemma 4.4.1, we have:

$$\mathbf{R}(\mathbf{N}_n ; n = 1, 2, \cdots) = \mathbf{R}(\mathbf{N}_{m'} ; m = 1, 2, \cdots) = \mathbf{M} \tag{4.4.θ}$$

(4.4.η), (4.4.θ) and our other observations establish that $\mathbf{M}$ is approximately finite of type $[p_1, p_2, \cdots]$ as stated.

§4.5 We proceed now to the second operation announced at the beginning of §4.3.

DEFINITION 4.5.1. *A ring $\mathbf{N}$ is of finite order if it possesses a finite linear basis, i.e. a (fixed finite) system, $A^0_1, \cdots, A^0_q$ such that $\mathbf{N}$ is the set of all*

$$A = \textstyle\sum_{u=1}^{q} x_u A^0_u \qquad (x_1, \cdots, x_q \textit{ any complex numbers}).$$

DEFINITION 4.5.2. $\mathbf{M}$ *is approximately finite (B) if this is true:*

Given any $A_1, \cdots, A_m \in \mathbf{M}$ and any $\epsilon > 0$ there exists a ring $\mathbf{N}$ with the following properties

(i) *$\mathbf{N}$ is of finite order*

(ii) $\mathbf{N} \subseteq \mathbf{M}$

(iii) *There exists $B_1, \cdots, B_m \in \mathbf{N}$ with*

$$[[B_h - A_h]] < \epsilon \quad \textit{for} \quad h = 1, \cdots, m.$$

LEMMA 4.5.1. *If $\mathbf{M}$ is approximately finite of type $[p_1, p_2, \cdots]$ then it is also approximately finite (B).*

PROOF. Suppose $A_1, \cdots, A_m \in \mathbf{M}$ and a $\epsilon > 0$ are given. Form the

$\mathbf{N}_1, \mathbf{N}_2, \cdots$ of Def. 4.1.1. By (ii) in Lemma 4.1.4, there exists for each $h = 1, \cdots, m$ an $\mathbf{N}_{n_h}$ such that for some $B_h \epsilon \mathbf{N}_{n_h}$

$$(4.5.\alpha) \qquad [[B_h - A_h]] < \epsilon.$$

Put $n = \text{Max}(n_1, \cdots, n_m)$ then $B_h \epsilon \mathbf{N}_{n_h} \subseteq \mathbf{N}_n$ i.e.

$$(4.5.\beta) \qquad B_1, \cdots, B_m \epsilon \mathbf{N}_n.$$

Now $\mathbf{N}_n$ is a factor in the case (I_{p_n}). By (ii) in Lemma 4.1.1 it is a ring of finite order. This, together with (4.5.α) and (4.5.β), completes the proof.

Comparison of Lemmas 4.4.3 and 4.5.1 shows that if we can derive approximate finiteness (A) from (B), then the equivalence of all kinds of approximate finiteness is established. This we accomplish in the next section. In the present section we obtain certain preliminary lemmas on rings of finite order; indeed we shall determine the structure of these.[37] Thus for the present section $\mathbf{N}$ is to be a ring of finite order.

LEMMA 4.5.2. (i) *There is a maximum r such that there is a system* $E_1, \cdots, E_r$ *of mutually orthogonal projections* $\neq 0$ *and* $\epsilon \mathbf{N}$.

(ii) *For such a system* (*with r maximum*) $E = \sum_{s=1}^{r} E_s$ *is the unit of* $\mathbf{N}$.

(iii) *Every* E_s *in such a system* (*with r maximum*) *is minimal* (*with respect to* $\mathbf{N}$) *in the sense of* [5], *p.* 143, *Def.* 5.1.2.

PROOF. Ad (i): Any such system is necessarily linearly independent. Hence the number of its elements must be $\leq$ the q of Def. 4.5.1. Hence the numbers for which there is such a system have a maximum.

Ad (ii): Let E be the unit of $\mathbf{N}$. $\sum_{s=1}^{r} E_s$ is a projection in $\mathbf{N}$; hence $\sum_{s=1}^{r} E_s \leq E$. So $E_{r+1} = E - \sum_{s=1}^{r} E_s$ is a projection in $\mathbf{N}$ which is orthogonal to the original E_s. Hence the assumed maximum property of r necessitates $E_{r+1} = 0$, i.e. $E = \sum_{s=1}^{r} E_s$.

Ad (iii): Consider a projection $F \leq E_s$ in $\mathbf{N}$. Then $E_s' = F$ and $E_s'' = E_s - F$ are mutually orthogonal projections in $\mathbf{N}$ which are also orthogonal to all $E_{s'}$, $s' \neq s$. Hence the assumed maximum property of r necessitates $E_s' = 0$ or $E_s'' = 0$ i.e. $F = 0$ or E_s. This however is the defining property of minimality.

REMARK. The center $\mathbf{N} \cdot \mathbf{N}'$ of $\mathbf{N}$ is of finite order along with $\mathbf{N}$. Hence we can apply the above lemma to it, thus obtaining a system $\bar{E}_1, \cdots, \bar{E}_{\bar{r}}$.

By this lemma, $\bar{E} = \sum_{s=1}^{\bar{r}} \bar{E}_s$ is the unit of $\mathbf{N} \cdot \mathbf{N}'$, hence by [2], p. 393, Theorems 3 and 4, it is also the unit of $\mathbf{N}$.

LEMMA 4.5.3. (i) $\bar{E}_s$ (*cf. above*) *is such that its range* $\mathfrak{M}_s$ *reduces* $\mathbf{N}, \mathbf{N}', \mathbf{N} \cdot \mathbf{N}'$

$$\text{(ii)} \qquad (\mathbf{N} \cdot \mathbf{N}')_{(\mathfrak{M}_s)} = (\bar{E}_s)_{(\mathfrak{M}_s)}$$

PROOF. Ad (i): This is implied by $\bar{E}_s \epsilon \mathbf{N} \cdot \mathbf{N}'$.

Ad (ii): This is equivalent to the statement $\bar{E}_s A = A\bar{E}_s = \alpha \bar{E}_s$ for all $A \epsilon \mathbf{N} \cdot \mathbf{N}'$. Put $\bar{E}_s A = A\bar{E}_s = A'$. Clearly $A' \epsilon \mathbf{N} \cdot \mathbf{N}'$, $\bar{E}_s A' = A'\bar{E}_s = A'$. Since $\bar{E}_s$ is a minimal projection in $\mathbf{N} \cdot \mathbf{N}'$ this implies $A' = \alpha \bar{E}_s$ (cf. the last part of the proof of Lemma 5.1.3 in [5], p. 144).

LEMMA 4.5.4. *$\mathbf{N}_{(\mathfrak{M}_s)}$ is a factor (in $\mathfrak{M}_s$) in a case* (I_{q_s}) $(q_s = 1, 2, \cdots)$.

PROOF. If A' is in the center of $\mathbf{N}_{(\mathfrak{M}_s)}$, $A'\bar{E}_s$ defines an operator A in $\mathfrak{H}$ which is readily seen to be in the center of $\mathbf{N}$ and for which $A\bar{E}_s = \bar{E}_s A = A$. Thus the center of $\mathbf{N}_{(\mathfrak{M}_s)}$ is included in $(\mathbf{N}\cdot\mathbf{N}')_{(\mathfrak{M}_s)}$. The converse inclusion is obvious.

It follows that the center of $\mathbf{N}_{(\mathfrak{M}_s)}$ is $(\mathbf{N}\cdot\mathbf{N}')_{(\mathfrak{M}_s)}$, and hence $\mathbf{N}_{(\mathfrak{M}_s)}$ is a factor by (ii) in Lemma 4.5.3.

$\mathbf{N}_{(\mathfrak{M}_s)}$ is of finite order along with $\mathbf{N}$. Therefore Lemma 4.5.2 implies that $\mathbf{N}_{(\mathfrak{M}_s)}$ contains a minimal projection. Thus $\mathbf{N}_{(\mathfrak{M}_s)}$ is a direct factor (cf. [5], pp. 150, 173, Theorem IV and Lemma 8.6.1). Obviously, since it is of finite order, $\mathbf{N}_{(\mathfrak{M}_s)}$ is not in a case (I_∞), and hence it is in a case (I_{q_s}), $q_s = 1, 2, \cdots$. This completes the proof.

LEMMA 4.5.5. *$\mathbf{N}$ possesses a finite linear base which consists of the operators $W^s_{u,v}$, $s = 1, \cdots, \bar{r}$, $u, v = 1, \cdots, q_s$ where for any fixed s the $W^s_{u,v}$, $u, v = 1, \cdots, q_s$ are a system of q_s-order matrix units with $\bar{E}_s = \sum_{u=1}^{q_s} W^s_{u,u}$. Thus for the q of Def. 4.5.1, $q = \sum_{s=1}^{\bar{r}} (q_s)^2$.*

PROOF. In virtue of Lemma 4.5.4, we may apply Lemma 4.1.1 and Remark 2 after it to $\mathbf{N}_{(\mathfrak{M}_s)}$. So $\mathbf{N}_{(\mathfrak{M}_s)}$ possesses a finite linear base of operators $\tilde{W}^s_{u,v}$ (in $\mathfrak{M}_s$), $u, v = 1, \cdots, q_s$ which form a system of q_s-order matrix units in $\mathfrak{M}_s$ with $\bar{E}_{s(\mathfrak{M}_s)} = \sum_{u=1}^{q_s} \tilde{W}^s_{u,u}$.

Since $\tilde{W}^s_{u,v} \in \mathbf{N}_{(\mathfrak{M}_s)}$ there is a $W^{s0}_{u,v}$ in $\mathbf{N}$ for which the part in $\mathfrak{M}_s$ is $\tilde{W}^s_{u,v}$. Let $W^s_{u,v} = W^{s0}_{u,v}\bar{E}_s$. Since $\bar{E}_s \in \mathbf{N}\cdot\mathbf{N}'$, we have $W^s_{u,v} \in \mathbf{N}$ and also $W^s_{u,v} = W^s_{u,v}\bar{E}_s = \bar{E}_s W^s_{u,v}$. Clearly $(W^s_{u,v})_{(\mathfrak{M}_s)} = \tilde{W}^s_{u,v}$ and one can readily show that the $W^s_{u,v}$ form a system of q_s-order matrix units in $\mathfrak{H}$ with $\bar{E}_s = \sum_{u=1}^{q_s} W^s_{u,u}$.

Consider an $A_s \in \mathbf{N}$ with $\bar{E}_s A_s = A_s \bar{E}_s = A_s$. Then A_s has the part $A_{s(\mathfrak{M}_s)}$ in $\mathfrak{M}_s$ and 0 in $\mathfrak{M}_s^{\perp}$. In the sense then that $\bar{E}_s$ is a transformation from $\mathfrak{H}$ to $\mathfrak{M}_s$, $A_s = A_{s(\mathfrak{M}_s)}\bar{E}_s$ and also $W^s_{u,v} = \tilde{W}^s_{u,v}\bar{E}_s$. As $A_{s(\mathfrak{M}_s)} \in \mathbf{N}_{(\mathfrak{M}_s)}$, we have, from the first paragraph of the proof, that $A_{s(\mathfrak{M}_s)} = \sum_{u=1}^{q_s} \sum_{v=1}^{q_s} a^s_{u,v}\tilde{W}^s_{u,v}$ and hence $A_s = A_{s(\mathfrak{M}_s)}\bar{E}_s = \sum_{u=1}^{q_s} \sum_{v=1}^{q_s} a^s_{u,v}\tilde{W}^s_{u,v}\bar{E}_s = \sum_{u=1}^{q_s} \sum_{v=1}^{q_s} a^s_{u,v} W^s_{u,v}$.

Consider now any $A \in \mathbf{N}$. Put $A_s = \bar{E}_s A$ which also equals $A\bar{E}_s$ since $\bar{E}_s \in \mathbf{N}\cdot\mathbf{N}'$. By the above, $A_s = \sum_{u=1}^{q_s} \sum_{v=1}^{q_s} a^s_{u,v} W^s_{u,v}$. By the remark after Lemma 4.5.2, we have

$$A = \bar{E}A = \sum_{s=1}^{\bar{r}} \bar{E}_s A = \sum_{s=1}^{\bar{r}} \sum_{u=1}^{q_s} \sum_{v=1}^{q_s} a^s_{u,v} W^s_{u,v}.$$

Thus the $W^s_{u,v}$, $s = 1, \cdots, \bar{r}$, $u, v = 1, \cdots, q_s$ form indeed a finite linear basis of $\mathbf{N}$.

We now prove a lemma which possesses a certain interest of its own. It applies to any ring of finite order $\mathbf{N} \subseteq \mathbf{M}$. We assume that a representation of $\mathbf{N}$ in the sense of Lemma 4.5.5 is given, with the $W^s_{u,v}$ as described there.

---

[17] The four lemmas which follow are merely an *ad hoc* derivation of Wedderburn's structure theorems for semi-simple algebras—specialized to the present situation. They can also be interpreted as a summary of certain results of unitary representation theory including the theorem of Burnside-Frobenius-Schur for that case.

LEMMA 4.5.6. *Given a (fixed) $q = 1, 2, \cdots$ there exists a factor* $\mathbf{O}$ *in case* (I_q) *with* $\mathbf{N} \subseteq \mathbf{O} \subseteq \mathbf{M}$ *if and only if this is true for* $\mathbf{N}$:

$$(4.5.\gamma) \qquad \text{All } qD_{\mathbf{M}}(W^s_{1,1}) \text{ are integers.}$$

PROOF. Assume that an $\mathbf{O}$ with the specified properties exists.

The considerations at the beginning of the proof of (i) in Lemma 4.1.2 show again that $D_{\mathbf{M}}(G) = D_{\mathbf{O}}(G)$ for all projections $G \in \mathbf{O}(\subseteq \mathbf{M})$. Hence in particular, $D_{\mathbf{M}}(W^s_{1,1}) = D_{\mathbf{O}}(W^s_{1,1})$, since $W^s_{1,1}$ is a projection $\in \mathbf{N} \subseteq \mathbf{O}$. But as $\mathbf{O}$ is in case (I_q), $D_{\mathbf{M}}(W^s_{1,1}) = D_{\mathbf{O}}(W^s_{1,1})$ must have a value $(0, 1/q, \cdots, 1)$ i.e. $q \cdot D_{\mathbf{M}}(W^s_{1,1})$ must be an integer. Thus $(4.5.\gamma)$ holds.

Sufficiency: Assume that $\mathbf{N}$ fulfills $(4.5.\gamma)$. Considering Lemma 4.5.5, we must prove this:

$(4.5.\delta)$ $\begin{cases}\text{Let a system of mutually orthogonal projections } \neq 0 \text{ from } \mathbf{M} \text{ be given:} \\ \bar{E}_1, \cdots, \bar{E}_r. \text{ For each } s, \text{ let a system of } q_s\text{-order matrix units } W^s_{u,v} \\ \text{from } \mathbf{M} \text{ be given, with } \bar{E}_s = \sum_{u=1}^{q_s} W^s_{u,u} \text{ and } q_s = 1, 2, \cdots. \text{ Let for} \\ \text{a given } q \text{ every } qD_{\mathbf{M}}(W^s_{1,1}) \text{ be an integer. Then there exists a factor} \\ \mathbf{O} \subseteq \mathbf{M} \text{ in a case } (\mathrm{I}_q) \text{ such that all } W^s_{u,v} \in \mathbf{O}.\end{cases}$

We therefore disregard $\mathbf{N}$ completely and deal with $(4.5.\delta)$ alone.

PROOF OF $(4.5.\delta)$. We have $D_{\mathbf{M}}(W^s_{1,1}) = p_s/q$, $p_s = 1, 2, \cdots$. The projections $W^s_{1,1}$ and $W^s_{u,u}$ are equivalent in virtue of the partially isometric $W^s_{u,1}$ (cf. Lemma 2.6.1). Hence

$$(4.5.\epsilon) \qquad D_{\mathbf{M}}(W^s_{u,u}) = p_s/q, \qquad D_{\mathbf{M}}(\bar{E}_s) = p_s q_s/q.$$

We may assume that

$$(4.5.\zeta) \qquad \sum_{s=1}^{\bar{r}} \bar{E}_s = 1.$$

Otherwise we could put $\bar{E}_{\bar{r}+1} = 1 - \sum_{s=1}^{\bar{r}} \bar{E}_s$. Then $\bar{E}_{\bar{r}+1}$ is a projection $\neq 0$. Also $qD_{\mathbf{M}}(\bar{E}_{\bar{r}+1}) = q - \sum_{s=1}^{\bar{r}} qD_{\mathbf{M}}(\bar{E}_s)$ is an integer. Thus we could add $\bar{E}_{\bar{r}+1}$ to the system $\bar{E}_1, \cdots, \bar{E}_{\bar{r}}$ with $q_{\bar{r}+1} = 1$ and $W^{\bar{r}+1}_{1,1} = \bar{E}_{\bar{r}+1}$.

By $(4.5.\epsilon)$ and $(4.5.\zeta)$, $\sum_{s=1}^{\bar{r}} p_s q_s = q$. We put $l_s = \sum_{t=1}^{s} p_t q_t$ for $s = 0, 1, \cdots, \bar{r}$. Then

$$(4.5.\eta) \qquad \begin{cases} l_s = l_{s-1} + p_s q_s & \text{for } s = 1, \cdots, \bar{r} \\ l_0 = 0, \qquad l_{\bar{r}} = q. \end{cases}$$

As $\mathbf{M}$ is in case (II_1), by Lemma 2.6.2 there exists a factor $\mathbf{O}^+ \subseteq \mathbf{M}$ in the case (I_q) with its matrix base $W^+_{o,w}$, $o, w = 1, \cdots, q_s$. Thus $W^+_{o,w}$ is a system of q-order matrix units, with $\sum_{u=1}^{q} W^+_{u,u} = 1$.

Put

$$(4.5.\theta) \qquad E^+_s = \sum_{u=l_{s-1}+1}^{l_s} W^+_{u,u} \qquad \text{for } s = 1, \cdots, \bar{r}$$

$$(4.5.\iota) \qquad \begin{cases} W^{+s}_{u,v} = \sum_{i=1}^{p_s} W^+_{l_{s-1}+(u-1)p_s+i,\, l_{s-1}+(v-1)p_s+i} \\ \qquad\qquad\qquad\qquad \text{for } s = 1, \cdots, \bar{r},\ u, v = 1, \cdots, q_s. \end{cases}$$

Then one verifies by a direct computation that the E_s^+ are mutually orthogonal projections $\neq 0$ with $\sum_{s=1}^{\bar{r}} E_s^+ = 1$ and (using Def. 2.6.1) that the $W_{u,v}^{+s}$ form (for any fixed s) a system of q_s-order matrix units with $\sum_{u=1}^{q_s} W_{u,u}^{+s} = E_s^+$. All these operators are in $\mathbf{M}$.

The $W_{u,u}^+$ are equivalent (due to the $W_{u,v}^+$) and since $\sum_{u=1}^{q} W_{u,u}^+ = 1$, $D_{\mathbf{M}}(W_{u,u}^+) = 1/q$. Hence by (4.5.$\iota$), $D_{\mathbf{M}}(W_{1,1}^{+s}) = p_s/q$ and by (4.5.ϵ), $D_{\mathbf{M}}(W_{1,1}^{+s}) = D_{\mathbf{M}}(W_{1,1}^s)$. Consequently there exists a partially isometric operator $V_s \in \mathbf{M}$ with the initial projection $W_{1,1}^{+s}$ and final projection $W_{1,1}^s$.

Now put

$$U = \sum_{s=1}^{\bar{r}} \sum_{u=1}^{q_s} W_{u,1}^s V_s W_{1,u}^{+s}.$$

Clearly $U \in \mathbf{M}$. One verifies, by direct computation, $U^*U = UU^* = 1$ which implies that U is unitary, and similarly that

$$(4.5.\kappa) \qquad UW_{u,v}^{+s} = W_{u,v}^s U \quad \text{or} \quad W_{u,v}^s = UW_{u,v}^+ U^{-1}.$$

$\mathbf{O} = U\mathbf{O}^+U^{-1}$ is a factor $\subseteq \mathbf{M}$ and in case (I_q) along with $\mathbf{O}^+$. By (4.5.ι), $W_{u,v}^{+s} \in \mathbf{O}^+$ and hence by (4.5.κ) $W_{u,v}^s = UW_{u,v}^{+s}U^{-1} \in U\mathbf{O}^+U^{-1} = \mathbf{O}$. Thus all $W_{u,v}^s \in \mathbf{O}$.

§4.6 We are now in a position to complete our proof of the relationships between the various kinds of approximate finiteness.

LEMMA 4.6.1. *Suppose* $\mathbf{N}$, $\mathbf{M}$, $\bar{E}_1, \cdots, \bar{E}_{\bar{r}}$ *and the* $W_{u,v}^s$ *are as in Lemma* 4.5.5 *with* $\bar{E} = \sum_{s=1}^{\bar{r}} \bar{E} = 1$. *For any* $\epsilon > 0$ *there exists an* $n_5 = n_5(\mathbf{N}, \epsilon)$ *such that for every* $q \geqq n_5$ *there exists a projection* $E = E(q, \mathbf{N}, \epsilon)$ *which is* $\in \mathbf{M}\cdot\mathbf{N}'$ *and has the following properties*:

(i) $D_{\mathbf{M}}(1 - E) < \epsilon$

(ii) *All* $qD_{\mathbf{M}}(EW_{1,1}^s)$ *are* $\neq 0$ *and integers.*

PROOF. Given a $q = 1, 2, \cdots$ we wish to choose for each $s = 1, \cdots, \bar{r}$ a $k_s = 1, 2, \cdots$ with

$$(4.6.\alpha) \qquad q(D_{\mathbf{M}}(W_{1,1}^s) - \epsilon/\textstyle\sum_{s=1}^{\bar{r}} q_s) < k_s \leqq qD_{\mathbf{M}}(W_{1,1}^s).$$

This is certainly feasible if $q\cdot D_{\mathbf{M}}(W_{1,1}^s) \geqq 1$ and $q\epsilon/\sum_{s=1}^{\bar{r}} q_s \geqq 1$ for $s = 1, \cdots, \bar{r}$. We assume therefore that $q \geqq n_5$ with

$$(4.6.\beta) \qquad n_5 = n_5(\mathbf{N}, \epsilon) = \mathrm{Max}\,(\textstyle\sum_{s=1}^{\bar{r}} q_s/\epsilon,\ 1/D_{\mathbf{M}}(W_{1,1}^s),\ s = 1, \cdots, \bar{r}).$$

Now (4.6.α) and $k_s \neq 0$ give

$$D_{\mathbf{M}}(W_{1,1}^s) - \epsilon/\textstyle\sum_{s=1}^{\bar{r}} q_s < k_s/q \leqq D_{\mathbf{M}}(W_{1,1}^s), \qquad k_s/q \neq 0.$$

As $\mathbf{M}$ is in case (II_1), this implies that we can find a projection $G^s \in \mathbf{M}$ with

$$(4.6.\gamma) \qquad D_{\mathbf{M}}(G^s) = k_s/q \neq 0$$

and

$$(4.6.\delta) \qquad G^s \leqq W_{1,1}^s$$

and that then

$$(4.6.\epsilon) \qquad D_{\mathbf{M}}(W_{1,1}^s - G_s) < \epsilon/\textstyle\sum_{s=1}^{\bar{r}} q_s.$$

$W^s_{u,1}$ is a partially isometric operator with initial projection $W^s_{1,1}$ and final projection $W^s_{u,u}$. Its mapping carries therefore the $G^s \leqq W^s_{1,1}$ into a projection $G^s_u \epsilon \mathbf{M}$ with

$$(4.6.\zeta) \qquad G^s_u \leqq W^s_{u,u}.$$

The same mapping carries (4.6.ϵ) into

$$(4.6.\eta) \qquad D_{\mathbf{M}}(W^s_{u,u} - G^s_u) < \epsilon/\textstyle\sum_{s=1}^{\bar{r}} q_s.$$

Finally $W^s_{u,v} \cdot W^s_{v,1} = W^s_{u,1}$ implies that the mapping of $W^s_{u,v}$ carries G^s_v into G^s_u. This may be written

$$(4.6.\theta) \qquad W^s_{u,v} G^s_v = G^s_u W^s_{u,v}.$$

For $s \neq t$ or $v \neq w$, $W^s_{v,v}$ the initial projection of $W^s_{u,v}$ is orthogonal to $W^t_{w,w}$ hence, *a fortiori*, to $G^t_w \leqq W^t_{w,w}$. So we have

$$(4.6.\iota) \qquad W^s_{u,v} G^t_w = 0 \qquad \text{for } s \neq t \text{ or } v \neq w.$$

$$(4.6.\kappa) \qquad G^t_w W^s_{u,v} = 0 \qquad \text{for } s \neq t \text{ or } u \neq w$$

Now put $E = \sum_{s=1}^{\bar{r}} \sum_{u=1}^{q_s} G^s_u$. Clearly $E \epsilon \mathbf{M}$ and since $\sum_{s=1}^{\bar{r}} \sum_{u=1}^{q_s} W^s_{u,u} = \sum_{s=1}^{\bar{r}} \bar{E}_s = 1$ we have $1 - E = \sum_{s=1}^{\bar{r}} \sum_{u=1}^{q_s} (W^s_{u,u} - G^s_u)$. Hence (4.6.$\eta$) yields

$$(4.6.\lambda) \qquad D_{\mathbf{M}}(1 - E) < \epsilon.$$

(4.6.θ), (4.6.ι) and (4.6.κ) give

$$W^s_{u,v} E = E W^s_{u,v} (= W^s_{u,v} G^s_v = G^s_u W^s_{u,v}).$$

So E commutes with $W^s_{u,v}$ and hence by Lemma 4.5.5 with all of $\mathbf{N}$. Since $E \epsilon \mathbf{M}$ this yields

$$(4.6.\mu) \qquad E \epsilon \mathbf{M} \cdot \mathbf{N}'.$$

Finally, direct computation shows that

$$(4.6.\nu) \qquad E W^s_{1,1} = G^s_1 = G^s.$$

Our assertions now follow immediately. (4.6.μ) gives the preliminary one, (4.6.λ) implies (i), and (4.6.ν) and (4.6.γ) yield (ii). Thus the lemma is proven.

It is now possible to take the final step.

LEMMA 4.6.2. *If* $\mathbf{M}$ *is approximately finite* (B), *then it is also approximately finite* (A).

PROOF. We must verify the criteria of Def. 4.3.1 for $\mathbf{M}$.

Let therefore a system $A_1, \cdots, A_m \epsilon \mathbf{M}$ and an $\epsilon > 0$ be given. Apply Def. 4.5.2 to these $A_1, \cdots, A_m$ and to $\epsilon/2$ thus obtaining the ring $\mathbf{N}$ and $B_1, \cdots, B_m \epsilon \mathbf{N}$ described there. Thus

$$(4.6.o) \qquad [[B_s - A_s]] < \epsilon/2.$$

$\mathbf{N}$ is a ring of finite order $\subseteq \mathbf{M}$. We form $\bar{E}_1, \cdots, \bar{E}_{\bar{r}}$ and the $W^s_{u,v}$ in the sense of Lemma 4.5.5 for this $\mathbf{N}$. To these and to

$$(4.6.\pi) \qquad \epsilon' = (\epsilon/2 \operatorname{Max}^{\cdot}(|||B_s|||;\ s = 1, \cdots, \bar{r}))^2$$

we apply Lemma 4.6.1, thus obtaining $n_5 = n_5(\mathbf{N}, \epsilon')$. For any $q \geqq n_5$ we form the projection $E = E(q, \mathbf{N}, \epsilon')$ of Lemma 4.6.1.

Accordingly we have $E \in \mathbf{M}' \cdot \mathbf{N}$ and $\mathbf{N}_E = E \cdot \mathbf{N} = \mathbf{N}E$ is a finite order ring $\subseteq \mathbf{M}$ along with $\mathbf{N}$. Thus $E\bar{E}_1, \cdots, E\bar{E}_{\bar{r}}$ and the $EW^s_{u,v}$ perform for $\mathbf{N}_E$ the functions of $\bar{E}_1, \cdots, \bar{E}_{\bar{r}}$ and the $W^s_{u,v}$ for $\mathbf{N}$ (cf. Lemma 4.5.5). By (ii) in Lemma 4.6.1 all $qD_{\mathbf{M}}(EW^s_{1,1})$ are integers. Thus Lemma 4.5.6 is applicable to $\mathbf{N}_E$ and we obtain a factor $\mathbf{O}$ in case (I_q) with

$$(4.6.\rho) \qquad \mathbf{N}_E \subseteq \mathbf{O} \subseteq \mathbf{M}.$$

$B_s \in \mathbf{N}$ implies

$$(4.6.\sigma) \qquad C_s = EB_s \in \mathbf{N}_E \subseteq \mathbf{O}.$$

Now (i) in Lemma 4.6.1 gives $D_{\mathbf{M}}(1 - E) < \epsilon'$. Hence $[[1 - E]] < \epsilon'^{\frac{1}{2}}$[35] and therefore

$$[[B_s - C_s]] = [[B_s(1 - E)]] < |||B_s|||\,\epsilon'^{\frac{1}{2}}.$$

Hence by $(4.6.\pi)$

$$(4.6.\tau) \qquad [[B_s - C_s]] < \epsilon/2.$$

Combining $(4.6.o)$ and $(4.6.\tau)$ gives

$$(4.6.\upsilon) \qquad [[C_s - A_s]] < \epsilon.$$

Thus we can satisfy Def. 4.3.1 by putting $n_1 = n_1(A_1, \cdots, A_m; \epsilon) = n_5(\mathbf{N}, \epsilon)$[38] and then for $q \geqq n_1 = n_5$, $\mathbf{N}_1 = \mathbf{N}_1(q, A_1, \cdots, A_m, \epsilon) = \mathbf{O}$.[38] Then our $C_1, \cdots, C_m$ replace the $B_1, \cdots, B_m$ of Def. 4.3.1, and (i)–(iii) there obtain as follows: (i) holds by virtue of the construction of $\mathbf{O}$ (ii) follows from $(4.6.\rho)$, and (iii) coincides with $(4.6.\upsilon)$.

Thus the proof is completed.

We could now make the final steps as indicated after Lemma 4.5.1 above, but we introduce one more variant of the concept of approximate finiteness:

Definition 4.6.1. $\mathbf{M}$ *is approximately finite* (*C*) *if*:

There exists a sequence of rings $\mathbf{N}_1, \mathbf{N}_2, \cdots$ *with the following properties*:

(i) $\mathbf{N}_n$ *is of finite order*

(ii) $\mathbf{N}_1 \subseteq \mathbf{N}_2 \subseteq \cdots$

(iii) *The ring* $\mathbf{M}$ *is generated by the* $\mathbf{N}_1, \mathbf{N}_2, \cdots$

$$\mathbf{M} = \mathbf{R}(\mathbf{N}_n;\ n = 1, 2, \cdots).$$

---

[38] Observe that the (indirectly defined) expression at the extreme right has the proper dependence—i.e. that one indicated by the expression in the middle.

LEMMA 4.6.3. *If* **M** *is approximately finite of type* $[p_1, p_2, \cdots]$ *then it is also approximately finite* (C).

PROOF. Immediate by comparison of Defs. 4.1.1 and 4.6.1.

LEMMA 4.6.4. *If* **M** *is approximately finite* (C), *then it is also approximately finite* (B).

PROOF. (ii) and (iii) in Def. 4.6.1 coincide with those in Def. 4.1.1. Hence Lemma 4.1.4 applies to our $\mathbf{N}_1, \mathbf{N}_2, \cdots$.

Let then the $A_1, \cdots, A_m \in \mathbf{M}$ and the $\epsilon > 0$ of Def. 4.5.2 be given. Owing to Lemma 4.1.4, A_k is the limit of a metrically convergent sequence from the set-theoretical sum of the $\mathbf{N}_1, \mathbf{N}_2, \cdots$. Hence there exists a $B_k \in \mathbf{N}_{n_k}$ with

$$[[B_k - A_k]] < \epsilon.$$

Now put $n = \text{Max}(n_1, \cdots, n_m)$. Then every $B_k \in \mathbf{N}_{n_k} \subseteq \mathbf{N}_n$. Thus $\mathbf{N} = \mathbf{N}_n$ meets all the requirements (i)–(iii) of Def. 4.5.2.

We can now prove

THEOREM XII. *All kinds of approximate finiteness are equivalent to each other*: (A), (B), (C) *and every type* $[p_1, p_2, \cdots]$ (*for every sequence* $[p_1, p_2, \cdots]$ *which fulfills* (i), (ii) *of Lemma* 4.1.3).

We shall therefore use the expression *approximately finite* from now on, without any further qualification.

PROOF. The implications

$$(A) \rightarrow \text{type } [p_1, p_2, \cdots] \rightarrow (B) \rightarrow (A)$$

were established in Lemmas 4.4.3, 4.5.1 and 4.6.2 respectively. Consequently

$$\text{type } [p_1, p_2, \cdots] \leftrightarrows (A) \leftrightarrows (B).$$

Further

$$\text{type } [p_1, p_2, \cdots] \rightarrow (C) \rightarrow (B)$$

by Lemmas 4.6.3, 4.6.4 respectively; hence

$$\text{type } [p_1, p_2, \cdots] \leftrightarrows (C)$$

too.

§4.7 We have not shown so far that approximately finite factors exist at all. In §§5.2, 5.3 and 5.6 we shall give various examples of such factors. Actually it is more difficult to show that there are factors in case (II_1) which are not approximately finite.

Our immediate goal is to prove a general theorem which yields the existence of approximately finite factors as a byproduct.

A preparatory lemma is necessary:

LEMMA 4.7.1. *Let a factor* $\mathbf{N} \subseteq \mathbf{M}$ *in a case* (I_p), $p = 1, 2, \cdots$ *be given. If an* $A \in \mathbf{N}$ *possesses these two properties*

(i) $[[AX - XA]] \leq \epsilon \, ||| X |||$ *for every* $X \in \mathbf{N}$

(ii) $Tr_{\mathbf{M}}(A) = 0$,

then $[[A]] \leq \epsilon$.

Proof. For any unitary $U \in \mathbf{N}$ we have $||| U ||| = ||| U^{-1} ||| = 1$. Hence by (i),

$$[[UAU^{-1} - A]] = [[U(AU^{-1} - U^{-1})A]] \leqq [[AU^{-1} - U^{-1}A]] \leqq \epsilon.$$

Also by (ii)

$$Tr_{\mathbf{M}}(UAU^{-1}) = Tr_{\mathbf{M}}(A) = 0.$$

If $U_1, \cdots, U_m$ are unitary and $\in \mathbf{N}$ then we have by the above $[[U_iAU_i^{-1} - A]] \leqq \epsilon$, $Tr_{\mathbf{M}}(U_iAU_i^{-1}) = 0$. So for

$$(4.7.\alpha) \qquad B = (1/m)(\textstyle\sum_{i=1}^{m} U_iAU_i^{-1})$$

the evaluations

$$(4.7.\beta) \qquad [[B - A]] \leqq \epsilon$$

$$(4.7.\gamma) \qquad Tr_{\mathbf{M}}(B) = 0$$

obtain.

Now choose an isomorphism $\mathcal{K}$ between $\mathbf{N}$ and the system of all (complex) p-order matrices, in the sense of (α) in Lemma 4.1.1. Put

$$\mathcal{K}A = a = \{a_{t,s}\}, \qquad \mathcal{K}B = b = \{b_{t,s}\}.$$

Form next the operators $V, X_1, \cdots, X_p \in \mathbf{N}$ with

$$\mathcal{K}V = v = \{v_{s,t}\},\ v_{s,t} \begin{cases} = 1 & \text{if } s - t \equiv 1 \bmod p \\ = 0 & \text{otherwise} \end{cases}$$

$$\mathcal{K}X = x_r = \{x^r_{s,t}\},\ x^r_{s,t} \begin{cases} = 1 & \text{if } t = s \neq r \\ = -1 & \text{if } t = s = r \\ = 0 & \text{otherwise.} \end{cases}$$

The matrices v, x_r are unitary. Hence the operators V, X_r are also unitary.

Now put $m = p2^p$ and let $U_1, \cdots, U_m$ be the operators $V^hX_1^{k_1} \cdot \cdots \cdot X_p^{k_p}$, $h = 0, 1, \cdots, p - 1; k_1, \cdots, k_p = 0, 1$. Then a combination of these formulas gives, due to $(4.7.\alpha)$ by direct computation,[39]

$$b_{t,s} \begin{cases} = (1/p) \sum_{u=1}^{p} a_{u,u} = a_0 & \text{for } t = s \\ = 0 & \text{otherwise.} \end{cases}$$

Consequently

$$(4.7.\delta) \qquad B = a_0 \cdot 1$$

[39] This computation may be outlined as follows. Let $\epsilon^{(r)}_{s,t} = (-1)^{\delta_{r,s}+\delta_{r,t}}$, i.e.

$$\epsilon^{(r)}_{s,t} \begin{cases} = 1 \text{ if } r \neq t \text{ and } s \neq r \text{ or if } r = s = t \\ = -1 \text{ if either } r = s \text{ and } r \neq t \text{ or } r \neq s \text{ and } r = t \end{cases}$$

Then if $\mathcal{K}X_rAX_r^{-1} = c^{(r)} = (c^{(r)}_{s,t})$

$$c^{(r)}_{s,t} = \epsilon^r_{s,t}\, a_{s,t}.$$

(continued)

(4.7.δ) and (4.7.γ) give together $0 = Tr_{\mathbf{M}}(B) = a_0$, i.e. $B = 0$. Hence (4.7.β) becomes $[[A]] \leqq \epsilon$ as desired.

We are able to prove:

THEOREM XIII. *If* $\mathbf{M}$ *is a factor in case* (II_1), $\mathbf{M}$ *contains a subring* $\mathbf{O}$ *which is a factor in case* (II_1), *and approximately finite.*

PROOF. Choose a sequence $[p_1, p_2, \cdots]$ which fulfills (i), (ii) in Lemma 4.1.3, e.g. $p_n = 2^n$.

Put $p_0 = 1$ and $\mathbf{N}_0 = (\alpha 1)$. This is a factor $\subseteq \mathbf{M}$ and in case (I_1). We shall choose by induction for every $n = 1, 2, \cdots$ a factor $\mathbf{N}_n \subseteq \mathbf{M}$ in the case (I_{p_n}). If for any $n = 1, 2, \cdots, \mathbf{N}_{n-1}$ has already been chosen, we choose $\mathbf{N}_n$ in the following way: $\mathbf{N}_{n-1}$, $\mathbf{M}$ are factors in the cases ($\mathrm{I}_{p_{n-1}}$), (II_1) respectively; $\mathbf{N}_{n-1} \subseteq \mathbf{M}$, p_{n-1} is a divisor of p_n. Apply therefore Lemma 4.4.2 to $\mathbf{N}_{n-1}$, $\mathbf{M}$, p_{n-1}, p_n in place of its $\mathbf{N}$, $\mathbf{O}$, p, r. Put $\mathbf{N}_n = \mathbf{R}$. Then $\mathbf{N}_n$ is a factor $\subseteq \mathbf{M}$ in the case (I_{p_n}).

Thus

$$\mathbf{N}_1 \subseteq \mathbf{N}_2 \subseteq \cdots \subseteq \mathbf{M}. \tag{4.7.ϵ}$$

Denote by $\mathbf{S}$ the set-theoretical sum of the sequence $\mathbf{N}_1, \mathbf{N}_2, \cdots$. Owing to (4.7.$\epsilon$), $\mathbf{S}$ is an algebra along with $\mathbf{N}_1, \mathbf{N}_2, \cdots$.

Denote the set of the limits of all metrically convergent sequences from $\mathbf{S}$ by $\mathbf{S}_1$. The same argument as in the proof of Lemma 4.1.4 shows that $\mathbf{S}_1$ is an algebra along with $\mathbf{S}$ and then that it is a ring. Hence $\mathbf{S} \subseteq \mathbf{S}_1$ implies $\mathbf{R}(\mathbf{S}) \subseteq \mathbf{S}_1$. On the other hand $\mathbf{R}(\mathbf{S})$ is closed in the sense of metric convergence by Theorem I. Hence $\mathbf{S} \subseteq \mathbf{R}(\mathbf{S})$ implies $\mathbf{S}_1 \subseteq \mathbf{R}(\mathbf{S})$. Consequently $\mathbf{R}(\mathbf{S}) = \mathbf{S}_1$ and so

$$\mathbf{M} \supseteqq \mathbf{R}(\mathbf{N}_n, n = 1, 2, \cdots) = \mathbf{R}(\mathbf{S}) = \mathbf{S}_1. \tag{4.7.ζ}$$

Next we prove that the ring $\mathbf{S}_1$ is a factor, i.e. that every $A \in \mathbf{S}_1 \cdot \mathbf{S}_1'$ belongs to $(\alpha \cdot 1)$.

Let $B_0 = \sum_{k_1, \cdots, k_p = 0,1} (X_1^{k_1} \cdot \cdots \cdot X_p^{k_p}) A (X_1^{k_1} \cdot \cdots \cdot X_{k_p}^{k_p})^{-1}$ and $\mathcal{K} B_0 = d = (d_{t,s})$. Then

$$\begin{aligned} d_{t,s} &= \sum_{k_1, \cdots, k_p = 0,1} (\epsilon_{s,t}^{(1)})^{k_1} \cdot \cdots \cdot (\epsilon_{s,t}^{(p)})^{k_p} a_{s,t} \\ &= ((\epsilon_{s,t}^{(1)})^0 + (\epsilon_{s,t}^{(1)})^1) \cdot \cdots \cdot ((\epsilon_{s,t}^{(p)})^0 + (\epsilon_{s,t}^{(p)})^1) a_{s,t} \\ &= (1 + \epsilon_{s,t}^{(1)}) \cdot \cdots \cdot (1 + \epsilon_{s,t}^{(p)}) a_{s,t}. \end{aligned}$$

Unless $s = t$, $\epsilon_{s,t}^{(r)} = -1$ and $d_{s,t} = 0$. If $s = t$, $d_{s,s} = 2^p a_{s,s}$. Thus B_0 is a diagonal matrix.

The effect of V on a diagonal matrix is just to permute the diagonal terms cyclicly. Hence if B and $b_{s,t}$ are as in the text

$$B = (1/p2^p) \sum_{h=0}^{p-1} V^h B_0 V^{-h}$$

and

$$\begin{aligned} b_{s,t} &= 0 \quad \text{if} \quad s \neq t \\ b_{s,s} &= (1/p) \sum_{u=1}^{p} a_{u,u}. \end{aligned}$$

Consider an $A \in \mathbf{S}_1 \cdot \mathbf{S}_1'$. Choose $\epsilon > 0$. As $A \in \mathbf{S}_1$ there exists a $B_\epsilon \in \mathbf{N}_{p_{n_\epsilon}}$ with

$$(4.7.\eta) \qquad [[B_\epsilon - A]] < \epsilon.$$

Put $C_\epsilon = B_\epsilon - Tr_{\mathbf{M}}(B_\epsilon)\cdot 1$. Then $C_\epsilon \in \mathbf{N}_{p_{n_\epsilon}}$ too, and

$$(4.7.\theta) \qquad Tr_{\mathbf{M}}(C_\epsilon) = 0.$$

Consider an $X \in \mathbf{N}_{p_{n_\epsilon}}$. As $A \in \mathbf{S}_1'$, $X \in \mathbf{S}_1$, so $AX - XA = 0$. Also, clearly $B_\epsilon X - XB_\epsilon = C_\epsilon X - XC_\epsilon$. Hence $C_\epsilon X - XC_\epsilon = B_\epsilon X - XB_\epsilon = (B_\epsilon - A)X - X(B_\epsilon - A)$ and so $[[C_\epsilon X - XC_\epsilon]] \leqq 2 \,|||\, X \,|||\, [[B_\epsilon - A]]$. This and (4.7.$\eta$) yield

$$(4.7.\iota) \qquad [[C_\epsilon X - XC_\epsilon]] \leqq 2 \,|||\, X \,|||\cdot \epsilon.$$

(4.7.θ) and (4.7.ι) permit us to apply Lemma 4.7.1 to C_ϵ, 2ϵ in place of its A, ϵ. Thus

$$[[C_\epsilon]] \leqq 2\epsilon.$$

Put $Tr_{\mathbf{M}}(B_\epsilon) = \alpha_\epsilon$. Then the above equation means

$$[[B_\epsilon - \alpha_\epsilon\cdot 1]] \leqq 2\epsilon.$$

Hence by (4.7.η)

$$(4.7.\kappa) \qquad [[A - \alpha_\epsilon\cdot 1]] < 3\epsilon.$$

As (4.7.κ) holds for all $\epsilon > 0$, it proves that A is the limit of a metrically convergent sequence from $(\alpha\cdot 1)$. As $(\alpha\cdot 1)$ is obviously closed in this sense, we have $A \in (\alpha\cdot 1)$ as desired.

Thus $\mathbf{S}_1$ is a factor. The considerations at the beginning of the proof of (i) in Lemma 4.1.2 show again that $D_{\mathbf{S}_1}(E) = D_{\mathbf{M}}(E)$ for all projections $E \in \mathbf{S}_1$. Hence $\mathbf{S}_1$ is in a finite case, along with $\mathbf{M}$. If $\mathbf{S}_1$ were in a case (I_q), $q = 1, 2, \cdots$, we could argue as follows: $\mathbf{N}_{p_n} \subseteq \mathbf{S}_1$ and $\mathbf{N}_{p_n}$ is a factor in case (I_{p_n}). Hence p_n is a divisor of q by (i) in Lemma 4.1.2. But this is impossible, since $p_n \to \infty$ as $n \to \infty$. Therefore $\mathbf{S}_1$ must be in the case (II_1).

Now (4.7.ϵ), (4.7.ζ) prove that $\mathbf{S}_1$ is approximately finite of type $[p_1, p_2, \cdots]$—i.e. approximately finite by Theorem XII.

Thus $\mathbf{O} = \mathbf{S}_1$ meets all our requirements.

And to conclude:

THEOREM XIV. *Approximately finite factors* $\mathbf{M}$ (*in the case* II_1) *exist and they have all the same algebraical type* $\overline{\mathbf{M}}$.

We denote this algebraical type by $\bar{\mathbf{A}}$.

PROOF. The existence follows from Theorem XIII.

Consider two approximately finite $\mathbf{M}_1$, $\mathbf{M}_2$. Choose a sequence $[p_1, p_2, \cdots]$ which fulfills (i), (ii) in Lemma 4.1.3, e.g. $p_n = 2^n$. Then $\mathbf{M}_1$, $\mathbf{M}_2$ are both approximately finite of type $[p_1, p_2, \cdots]$ by Theorem XII. Hence they are algebraically isomorphic by Theorem XI, i.e. $\overline{\mathbf{M}}_1 = \overline{\mathbf{M}}_2$.

Thus $\bar{\mathbf{M}}$ ($\mathbf{M}$ approximately finite) is uniquely determined and the definition of $\bar{\mathbf{A}}$ is justified.

§**4.8** The algebraical isomorphism invariants described at the end of §2.10 will now be studied for the algebraical type $\bar{\mathbf{A}}$. These invariants are

(1) Validity of $\bar{\mathbf{A}}^c = \bar{\mathbf{A}}$

(2) The fundamental group: $\mathfrak{G} = \mathfrak{G}(\bar{\mathbf{A}})$.

LEMMA 4.8.1. $\bar{\mathbf{A}}^c = \bar{\mathbf{A}}$.

PROOF. By Theorem XIV on one hand and Theorem III on the other, it suffices to show that a definition of approximate finiteness is unaffected by a conjugate or by a dual isomorphism. This is obviously true, even for all the definitions given in Theorem XII and for both types of general isomorphisms (cf. B^*) in §2.2) mentioned above.

LEMMA 4.8.2. $\mathfrak{G} = \mathfrak{G}(\bar{\mathbf{A}})$ *contains all rational numbers* θ *in* $0 < \theta < \infty$.

PROOF. Since $\mathfrak{G}$ is a group, it suffices to prove that it contains all $1/p$, $p = 2, 3, \cdots$.

Consider such a $1/p$. Choose a sequence $[p_1, p_2, \cdots]$ which fulfills (i), (ii) in Lemma 4.1.3 and with all p_m divisible by p, e.g. $p_n = p^n$.

Choose an approximately finite $\mathbf{M}$ and the $\mathbf{N}_1, \mathbf{N}_2, \cdots$, for this $\mathbf{M}$ and the sequence $[p_1, p_2, \cdots]$ in the sense of Def. 4.1.1.

The considerations at the beginning of the proof of (i) in Lemma 4.1.2 show again that $D_{\mathbf{M}}(G) = D_{\mathbf{N}_n}(G)$ for all projections $G \in \mathbf{N}_n (\subseteq \mathbf{M})$.

As $\mathbf{N}_1$ is in the case (I_{p_1}), p_1 divisible by p, there exists a projection $E \in \mathbf{N}_1$ with $D_{\mathbf{N}_1}(E) = 1/p$. Thus

$$E \in \mathbf{N}_1 \subseteq \mathbf{N}_2 \subseteq \cdots \subseteq \mathbf{M} \tag{4.8.α}$$

while $D_{\mathbf{N}_1}(E) = 1/p$ yields

$$D_{\mathbf{O}}(E)/D_{\mathbf{O}}(1) = D_{\mathbf{O}}(E) = 1/p \quad \text{for} \quad \mathbf{O} = \mathbf{N}_1, \mathbf{N}_2, \cdots, \mathbf{M}. \tag{4.8.β}$$

Denote the range of E by $\mathfrak{M}$.

By (4.8.α) we can form the factors $\mathbf{N}_{1(\mathfrak{M})}, \mathbf{N}_{2(\mathfrak{M})}, \cdots \mathbf{M}_{(\mathfrak{M})}$ in $\mathfrak{M}$, and we have

$$\mathbf{N}_{1(\mathfrak{M})} \subseteq \mathbf{N}_{2(\mathfrak{M})} \subseteq \cdots \subseteq \mathbf{M}_{(\mathfrak{M})}. \tag{4.8.γ}$$

By (4.8.β), $\overline{\mathbf{N}_{n(\mathfrak{M})}} = \bar{\mathbf{N}}_n^{1/p}$, $\overline{\mathbf{M}_{(\mathfrak{M})}} = \bar{\mathbf{M}}^{1/p}$ (cf. the discussion before Theorem VI). The former implies by Lemma 2.10.3 that

$$\mathbf{N}_{n(\mathfrak{M})} \text{ is in the case } (\mathrm{I}_{p_n/p}). \tag{4.8.δ}$$

We restate the latter:

$$\overline{\mathbf{M}_{(\mathfrak{M})}} = \bar{\mathbf{M}}^{1/p}. \tag{4.8.ϵ}$$

We know that $\mathbf{M}$ is the closure of the set-theoretical sum of $\mathbf{N}_1, \mathbf{N}_2, \cdots$ in the strong topology.[40] Hence it follows from (4.8.α) that $\mathbf{M}_{(\mathfrak{M})}$ is the closure of

[40] That closure is obviously $= \mathbf{R}(\mathbf{N}_n, n = 1, 2, \cdots)$. Hence it is $\mathbf{M}$ by (iii) in Def. 4.1.1.

the set-theoretical sum of $\mathbf{N}_{1(\mathfrak{M})}$, $\mathbf{N}_{2(\mathfrak{M})}$, $\cdots$ in the strong topology. This and (4.8.γ) give together

$$(4.8.\zeta) \qquad \mathbf{M}_{(\mathfrak{M})} = \mathbf{R}(\mathbf{N}^{*}_{n(\mathfrak{M})}, n = 1, 2, \cdots).$$

Now (4.8.δ), (4.8.γ) and (4.8.ζ) coincide with (i), (ii), (iii) in Def. 4.1.1 as applied to $\mathbf{M}_{(\mathfrak{M})}$, $\mathbf{N}_{1(\mathfrak{M})}$, $\mathbf{N}_{2(\mathfrak{M})}$, $\cdots$, p_1/p, p_2/p_2, $\cdots$ in place of its $\mathbf{M}$, $\mathbf{N}_1$, $\mathbf{N}_2 \cdots$, p_1, p_2, $\cdots$. Hence $\mathbf{M}_{(\mathfrak{M})}$ is approximately finite.

As $\mathbf{M}$ is approximately finite too, this shows that $\overline{\mathbf{M}_{(\mathfrak{M})}} = \overline{\mathbf{M}}$ by Theorem XIV. Owing to (4.8.ϵ), this gives $\overline{\mathbf{M}}^{1/p} = \overline{\mathbf{M}}$. Hence $1/p \in \mathfrak{G}$ as desired.

LEMMA 4.8.3. *If there exists for every $\epsilon > 0$ an α in $1 - \epsilon < \alpha < 1$ such that $\overline{\mathbf{M}}^{\alpha}$ is approximately finite, then $\mathbf{M}$ itself is approximately finite.*

PROOF. Assume that $\mathbf{M}$ possesses the property described above. We shall establish its approximate finiteness in the sense of Def. 4.5.2.

Let accordingly $A_1, \cdots, A_m \in \mathbf{M}$ and an $\epsilon > 0$ be given. We shall construct an $\mathbf{N}$ which fulfills (i)–(iii) in that definition.

Apply first our hypothesis concerning $\mathbf{M}$ to

$$(4.8.\eta) \qquad \epsilon' = (\epsilon/4 \operatorname{Max} (||| A_s |||, s = 1, \cdots, m))^2$$

in place of its ϵ. Choose the α mentioned there with $1 - \epsilon' < \alpha < 1$. Then choose a projection $E \in \mathbf{M}$ with $D_{\mathbf{M}}(E) = \alpha$. Let $\mathfrak{M}$ be the range of E. Form $\mathbf{M}_{(\mathfrak{M})}$. Then we have $\overline{\mathbf{M}_{(\mathfrak{M})}} = \overline{\mathbf{M}}^{\alpha}$ (cf. the discussion before Theorem VI). It follows from our assumptions concerning α that $\mathbf{M}_{(\mathfrak{M})}$ is approximately finite.

By [5], p. 187, (i), in Lemma 11.3.3,

$$A \leftrightarrows A_{(\mathfrak{M})}$$

is a one-to-one correspondence between all $A \in \mathbf{M}$ with $EA = AE = A$ and all $A_{(\mathfrak{M})} \in \mathbf{M}_{(\mathfrak{M})}$. Now we can repeat with a slight variation an argument that was used in the proof of Theorem XI: Clearly $cTr_{\mathbf{M}}(A)$ will do as $Tr_{\mathbf{M}_{(\mathfrak{M})}}(A_{(\mathfrak{M})})$ in the sense of [6], p. 219, Property IV, for the A, $A_{(\mathfrak{M})}$ described above, if we choose the normalizing factor $c > 0$ so that $cTr_{\mathbf{M}}(A) = 1$ when $A_{(\mathfrak{M})}$ is the unit of $\mathbf{M}_{(\mathfrak{M})}$. This means $A = E$, so we want $cTr_{\mathbf{M}}(E) = 1$. Since $Tr_{\mathbf{M}}(E) = D_{\mathbf{M}}(E) = \alpha$ this means $c = 1/\alpha$. And as the trace is uniquely characterized by the above property, $Tr_{\mathbf{M}_{(\mathfrak{M})}}(A_{(\mathfrak{M})}) = (1/\alpha)Tr_{\mathbf{M}}(A)$. From this $(1/\alpha)Tr_{\mathbf{M}}(A^*A) = Tr_{\mathbf{M}_{(\mathfrak{M})}}((A^*A)_{(\mathfrak{M})}) = Tr_{\mathbf{M}_{(\mathfrak{M})}}((A^*)_{(\mathfrak{M})} \cdot (A)_{(\mathfrak{M})})$ for the same A, $A_{(\mathfrak{M})}$ hence

$$(4.8.\theta) \qquad [[A_{(\mathfrak{M})}]]_{\mathbf{M}_{(\mathfrak{M})}} = (1/\alpha)^{\frac{1}{2}}[[A]]_{\mathbf{M}}.$$ [41]

Now put $B_s = EA_sE$, $s = 1, \cdots, m$. Then we have $D_{\mathbf{M}}(1 - E) < \epsilon'$. Hence $[[1 - E]]_{\mathbf{M}} < \epsilon'^{\frac{1}{2}}$ [35] and therefore $[[A_s - B_s]]_{\mathbf{M}} = [[A_s(1 - E) + (1 - E)A_sE]] < ||| A_s ||| \epsilon'$. Hence by (4.8.$\eta$)

$$(4.8.\iota) \qquad [[A_s - B_s]]_{\mathbf{M}} < \epsilon/2.$$

[41] As a rule we did not indicate the factor (in a finite case) with respect to which $[[A]]$ is formed. However, here it is desirable.

Now $B_s \in \mathbf{M}$ and $EB_s = B_sE = B_s$ imply $B_{s(\mathfrak{M})} \in \mathbf{M}_{(\mathfrak{M})}$. We now apply Def. 4.5.2 to the approximately finite $\mathbf{M}_{(\mathfrak{M})}$ with $B_{1(\mathfrak{M})}, \cdots, B_{m(\mathfrak{M})}$, $\epsilon/2\alpha^{\frac{1}{2}}$ in place of its $A_1, \cdots, A_m$ and ϵ. This yields a ring $\mathbf{N}^+ \subseteq \mathbf{M}_{(\mathfrak{M})}$ of finite order with $C_1^+, \cdots, C_m^+ \in \mathbf{N}^+$ such that

$$(4.8.\kappa) \qquad [[C_s^+ - B_{s(\mathfrak{M})}]]_{\mathbf{M}_{(\mathfrak{M})}} < \epsilon/2\alpha^{\frac{1}{2}}.$$

As $\mathbf{N}^+ \subseteq \mathbf{M}_{(\mathfrak{M})}$ we extend every $C^+ \in \mathbf{N}^+$ to a $C \in \mathbf{M}$ with $EC = CE = C$ and $C^+ = C_{(\mathfrak{M})}$.[42] These C form obviously a ring $\mathbf{N} \subseteq \mathbf{M}$ of finite order with $\mathbf{N}^+ = \mathbf{N}_{(\mathfrak{M})}$. By this process the $C_1^+, \cdots, C_m^+ \in \mathbf{N}^+$ give rise to the $C_1, \cdots, C_m \in \mathbf{N}$. Now (4.8.$\theta$) gives

$$[[C_s^+ - B_{s(\mathfrak{M})}]]_{\mathbf{M}_{(\mathfrak{M})}} = [[(C_s - B_s)_{(\mathfrak{M})}]]_{\mathbf{M}_{(\mathfrak{M})}} = (1/\alpha^{\frac{1}{2}})[[C_s - B_s]]_{\mathbf{M}}$$

and so (4.8.κ) becomes

$$(4.8.\lambda) \qquad [[C_s - B_s]]_{\mathbf{M}} < \epsilon/2.$$

(4.8.ι) and (4.8.λ) together give

$$(4.8.\mu) \qquad [[C_s - A_s]]_{\mathbf{M}} < \epsilon.$$

Thus the ring $\mathbf{N} \subseteq \mathbf{M}$ of finite order meets the requirements (i)–(iii) in Def. 4.5.2. Hence $\mathbf{M}$ is approximately finite as asserted.

LEMMA 4.8.4. *$\mathfrak{S} = \mathfrak{S}(\bar{\mathbf{A}})$ is the set P of all (real) numbers θ in $0 < \theta < \infty$.*

PROOF. Consider a θ in $0 < \theta < \infty$. By Lemma 4.8.2 we have for every rational ξ in $0 < \xi < \infty$, $\bar{\mathbf{A}}^{\xi} = \bar{\mathbf{A}}$. Hence $\bar{\mathbf{A}}^{\xi}$ is approximately finite. Therefore $(\bar{\mathbf{A}}^{\theta})^{\alpha} = \bar{\mathbf{A}}^{\theta\alpha}$ is approximately finite if $\theta\alpha$ is rational.

Now it is obviously possible to find for any given $\epsilon > 0$ an α in $1 - \epsilon < \alpha < 1$ for which $\theta\alpha$ is rational. This makes $(\bar{\mathbf{A}}^{\theta})^{\alpha}$ approximately finite. Hence $\bar{\mathbf{A}}^{\theta}$ itself is approximately finite by Lemma 4.8.3 with $\bar{\mathbf{A}}^{\theta}$ in place of its $\mathbf{M}$. Consequently Theorem XIV gives $\bar{\mathbf{A}}^{\theta} = \mathbf{A}$. Hence $\theta \in \mathfrak{S}$ as desired.

We restate the results of Lemmas 4.8.1 and 4.8.4.

THEOREM XV. *The algebraical isomorphism invariants in* (1), (2) *at the beginning of this section, behave for the approximate finite type $\bar{\mathbf{A}}$ as follows*:

(1) $\bar{\mathbf{A}}^{c} = \bar{\mathbf{A}}$

(2) $\mathfrak{S}(\bar{\mathbf{A}}) = P$.

It is worth pointing out that the two invariants (1), (2) are not sufficient to characterize an algebraical type $\overline{\mathbf{M}}$ in the case (II_1). A proof of this is only possible if we show first that not all $\overline{\mathbf{M}}$ are approximately finite: If all were, then Theorem XIV would give $\overline{\mathbf{M}} = \bar{\mathbf{A}}$, i.e. only one $\overline{\mathbf{M}}$ in case (II_1) would exist and thus no invariants for these $\overline{\mathbf{M}}$ would be needed at all.

Now we shall prove in Theorems XVI, XVI′ that not approximately finite $\overline{\mathbf{M}}$ in the case (II_1) exist. By making advance use of this result we prove:

LEMMA 4.8.5. *The algebraical isomorphism invariants in* (1), (2) *at the beginning of this paragraph are not sufficient to characterize the algebraical type $\overline{\mathbf{M}}$ in the case* (II_1).

[42] C is reduced by $\mathfrak{M}$, its part in $\mathfrak{M}$ is C^+ and its part in $\mathfrak{M}^+$ is 0.

PROOF. Assume the opposite.

Consider a type $\overline{\mathbf{M}}$ in the case (II_1).

Owing to the formulae (2.3.α) and (2.8.β) the validity of $\overline{\mathbf{M}}^c = \overline{\mathbf{M}}$ is unaffected if we replace $\overline{\mathbf{M}}$ by $\overline{\mathbf{M}}^c$ or by an $\overline{\mathbf{M}}^\theta (0 < \theta < \infty)$. By Lemma 2.10.2 $\mathfrak{G}(\overline{\mathbf{M}})$ is unchanged if we replace $\overline{\mathbf{M}}$ by $\overline{\mathbf{M}}^c$ or by an $\overline{\mathbf{M}}^\theta (0 < \theta < \infty)$. In other words: The invariants (1) and (2) are unaffected by a replacement of $\overline{\mathbf{M}}$ by $\overline{\mathbf{M}}^c$ or by an $\overline{\mathbf{M}}^\theta$.

Therefore our assumptions concerning these invariants permit us to conclude $\overline{\mathbf{M}} = \overline{\mathbf{M}}^c$ and $\overline{\mathbf{M}} = \overline{\mathbf{M}}^\theta$. The latter equations imply $\theta \in \mathfrak{G}(\overline{\mathbf{M}})$ and therefore $\mathfrak{G}(\overline{\mathbf{M}}) = P$.

Thus the two invariants (1) and (2) are uniquely determined, independently of the choice of $\overline{\mathbf{M}}$ in the case (II_1).

Hence a second application of our assumption concerning these invariants yields that there exists only one type $\overline{\mathbf{M}}$ in the case (II_1). $\bar{\mathbf{A}}$ is such a type; hence all $\overline{\mathbf{M}}$ in the case (II_1) must equal $\bar{\mathbf{A}}$, i.e. be approximately finite.

However, as stated above, the opposite will be proven in Theorems XVI, XVI′, and we propose to use this now. Thus we have a contradiction and the proof is completed.

We conclude with the observation that the existence of not approximately finite factors $\mathbf{M}$ in the case (II_1) will be established with the help of algebraic invariants other than (1), (2) above (cf. the beginning of §6.1).

CHAPTER V. ON VARIOUS EXAMPLES

§5.1 We established the existence of approximately finite factors (i.e. of their type $\bar{\mathbf{A}}$) by the indirect procedure of Theorems XIII and XIV.[43] It is therefore desirable to furnish examples of this by explicit and direct construction and also to investigate those examples of factors in the case (II_1) which we possess already as to their approximate finiteness. The latter investigation is also necessary in order to find not approximately finite factors in the case (II_1).

In connection with this work we shall develop a new technique to construct examples of factors in the case (II_1), which is a simplification of the original one of [5], pp. 192–209, Part IV.

We begin by establishing the approximate finiteness of some of our old examples: certain ones in [5] and that one in [7], pp. 72–77, §7.5.

§5.2 LEMMA 5.2.1. *The ring* $\mathbf{C}^{\#1}$ *of* [7], *p.* 72, *Lemma* 7.5.1 *is approximately finite.*

PROOF. Using the notations loc. cit., particularly pp. 68, 72, we have

$$\mathbf{C}^{\#1} = \mathbf{R}(\bar{\bar{\mathbf{B}}}_{(n,1)}\,;\, n = 1, 2, \cdots).$$

Hence

$$\mathbf{C}^{\#1} = \mathbf{R}(\mathbf{N}_m\,;\, m = 1, 2, \cdots) \tag{5.2.α}$$

[43] The decisive passage from Theorem XIII to Theorem XIV makes use of the existence of factors in the case (II_1), for instance the examples of [5].

with

$$(5.2.\beta) \qquad \mathbf{N}_m = \mathbf{R}(\bar{\bar{\mathbf{B}}}_{(n,1)}\,;\, n = 1, \cdots, m).$$

Thus $\mathbf{N}_1 \subseteq \mathbf{N}_2 \subseteq \cdots$ and therefore $\mathbf{C}^{\#1}$ is approximately finite by Def. 4.6.1 if the $\mathbf{N}_m$ are shown to be of finite order. [44]

These considerations belong in $\prod_{\otimes, n=1,2,\cdots}^{\mathfrak{D}} (\mathfrak{H}_{(n,1)} \otimes \mathfrak{H}_{(n,2)})$ where $\mathbf{C}^{\#1}$ is a factor in the case (II_1), but in establishing the finite order of $\mathbf{N}_m$, we may as well consider the entire space $\prod_{\otimes, n=1,2,\cdots} (\mathfrak{H}_{(n,1)} \otimes \mathfrak{H}_{(n,2)})$.

We proceed as in the lower part of p. 68, loc. cit. $\mathbf{B}_{(n,1)}$ is in $\mathfrak{H}_{(n,1)}$. It is first extended to $\bar{\mathbf{B}}_{(n,1)}$ in $\mathfrak{H}_{(n,1)} \otimes \mathfrak{H}_{(n,2)}$ and then to $\bar{\bar{\mathbf{B}}}_{(n,1)}$ in $\prod_{\otimes, n=1,2,\cdots} (\mathfrak{H}_{(n,1)} \otimes \mathfrak{H}_{(n,2)})$. Hence the $\bar{\bar{\mathbf{B}}}_{(m,1)}$ commute with each other. Therefore the $\mathbf{N}_m$ of (5.2.β) is of finite order if the $\bar{\bar{\mathbf{B}}}_{(n,1)}$ are. Now $\mathbf{B}_{(n,1)}$ is of finite order because it is in the finite 2-dimensional space $\mathfrak{H}_{(n,1)}$ and so $\bar{\bar{\mathbf{B}}}_{(n,1)}$ is of finite order too.

LEMMA 5.2.2. *The ring* $\mathbf{M}$ *of* [5], *pp.* 203–204, *Theorem XI, is approximately finite if the group* $\mathfrak{G}$ *possesses this property*:

(i) $\mathfrak{G}$ *is the set-theoretical sum of a sequence* $\mathfrak{G}_1 \subseteq \mathfrak{G}_2 \subseteq \cdots$ *of finite groups.*

(We assume, of course, that the requirements of loc. cit. [5], p. 206, Lemma 13.1.2, are fulfilled, placing $\mathbf{M}$ in the case (II_1).)

PROOF. Using the notations loc. cit., particularly those on pp. 198–200, we have

$$(5.2.\gamma) \qquad \mathbf{M} = \mathbf{R}(\bar{U}_{a_0}, \bar{L}_{\varphi(x)};\, a \in \mathfrak{G},\ \varphi(x) \text{ bounded and measurable}).$$

A familiar argument shows that the $\varphi(x)$ in (5.2.γ) can be restricted considerably without changing the ring $\mathbf{M}$. Indeed:

First. It suffices for reasons of continuity to consider the $\varphi(x)$ with finite ranges only.

Second. These $\varphi(x)$ are finite linear combinations of the characteristic functions

$$\chi_S(x) \begin{cases} = 1 & \text{for } x \in S \\ = 0 & \text{otherwise.} \end{cases}$$

Third. Let $T_1, T_2, \cdots$ be a countable basis for the T ($\subseteq S$ and measurable).[45]

[44] It is not difficult to show that $\mathbf{N}_n$ is a factor in the case (I_{2^n}). Hence Def. 4.1.1. would also be applicable, yielding approximate finiteness of type [2, 4, 8, $\cdots$].

[45] The space of all measurable sets T (disregarding sets of measure zero) can be metricized by the distance

$$\overline{T'T''} = \mu(T' + T'') - \mu(T'T'').$$

This is obviously equal to

$$|\chi_{T'} - \chi_{T''}|^2 = \int_S |\chi_{T'}(x) - \chi_{T''}(x)|^2\, dx.$$

Since $\mathfrak{H}_S$ is separable, there exists a sequence $\chi_{T_1}, \chi_{T_2}, \cdots$ dense in the set of all χ_T, and correspondingly the sequence $T_1, T_2, \cdots$ is dense in the set of all T. This is what we mean by a basis.

Then it suffices for reasons of continuity to consider these $\varphi(x) = \chi_{T_i}(x)$, $i = 1, 2, \cdots$ only.

Finally, these properties of the system $T_1, T_2, \cdots$ are not lost if we replace it by another countable system which contains it. We do this in such a way that the increased system $\bar{T}_1, \bar{T}_2, \cdots$ possesses these further properties:

(1) It contains $T' \cdot T''$ along with T', T''.

(2) It contains $a_0 T'$ along with T' for any $a_0 \in \mathfrak{G}$.

These being understood, we have

$$\mathbf{M} = \mathbf{R}(\bar{U}_{a_0}, \bar{L}_{\chi_{\bar{T}_i}(x)};\ a_0 \in \mathfrak{G}_1,\ i = 1, 2, \cdots). \tag{5.2.δ}$$

Consider now the finite groups $\mathfrak{G}_1, \mathfrak{G}_2, \cdots$ of (i). $\bar{T}_1, \cdots, \bar{T}_n$ generates, with the help of the operations (1) and (2) when a_0 in the latter is restricted to $\mathfrak{G}_n$, a finite subsystem of $\bar{T}_1, \bar{T}_2, \cdots$,—say the set $(\bar{T}_i, i \in I_n)$. So I_n is a finite set. Clearly $I_1 \subseteq I_2 \subseteq \cdots$. I_n contains $1, 2, \cdots, n$. Hence $(1, 2, \cdots)$ is the set-theoretical sum of $I_1, I_2, \cdots$.

Consequently (5.2.δ) gives

$$\mathbf{M} = \mathbf{R}(\mathbf{N}_n, n = 1, 2, \cdots) \tag{5.2.ϵ}$$

with

$$\mathbf{N}_n = \mathbf{R}(\bar{U}_{a_0}, \bar{L}_{\chi_{\bar{T}_i}(x)};\ a_0 \in \mathfrak{G}_n,\ i \in I_n). \tag{5.2.ζ}$$

Thus $\mathbf{N}_1 \subseteq \mathbf{N}_2 \subseteq \cdots$ and therefore $\mathbf{M}$ is approximately finite by Def. 4.6.1 if the $\mathbf{N}_n$ are shown to be of finite order.[46]

Now the group character of $\mathfrak{G}_n$ and the construction of I_n show that the operators

$$\bar{U}_{a_0}\bar{L}_{\chi_{\bar{T}_i}(x)}, \qquad a_0 \in \mathfrak{G}_n, \qquad i \in I_n \tag{5.2.η}$$

are reproduced by multiplication. As $\mathfrak{G}_n$, I_n are finite, these operators (5.2.η) form a finite set too. Hence they are a finite basis of $\mathbf{N}_n$ completing the proof.

This last result establishes the approximate finiteness of the examples (β), (γ) in [5], p. 208. Thus our Theorem XIV establishes their algebraic isomorphism and [6], p. 244, Theorem XI, their spatial isomorphism as announced by (v) in [5], p. 299. In order to include (α), p. 208, eod., we need this

LEMMA 5.2.3. *The condition (i) in Lemma 5.2.2 can be replaced by this condition:*

(ii) $\mathfrak{G}$ *is abelian.*

The proof of this lemma is somewhat complicated. It requires some rather deep results on the decompositions of mappings of measurable sets, which will be published elsewhere. We shall not pursue this matter further on this occasion.

§5.3 We proceed to expound the new (simplified) method for the construction of examples of factors in the case (II_1), announced in §5.1.

---

[46] This time we really must use Def. 4.6.1 and it would not be possible to replace it by Def. 4.1.1—in contrast with footnote[44].

Let a group $\mathfrak{G}$ be given. For the moment $\mathfrak{G}$ could be finite or countably infinite, but we shall be forced to assume the latter at a subsequent stage (cf. Lemma 5.3.4 and the Remark after it).

Form a unitary vector space $\mathfrak{H}$ by using the elements of $\mathfrak{G}$ as indices for the vector components, i.e. $\mathfrak{H}$ is the set of all complex vectors $f = [x_a\,;\, a \in \mathfrak{G}]$ with a finite $\sum_{a\in\mathfrak{G}} | x_a |^2$ and for $f = [x_a\,;\, a \in \mathfrak{G}]$ and $g = [y_a\,;\, a \in \mathfrak{G}]$, $(f, g) = \sum_{a\in\mathfrak{G}} x_a \bar{y}_a$ [47]

Now we introduce and discuss some operators in $\mathfrak{H}$.

LEMMA 5.3.1. *Define the operators*

$$\left.\begin{array}{lrcl} \text{(i)} & U_a[x_a\,;\, a \in \mathfrak{G}] & = & [x_{aa_0}\,;\, a \in \mathfrak{G}] \\ \text{(ii)} & V_{a_0}[x_a\,,\, a \in \mathfrak{G}] & = & [x_{a^{-1}a}\,;\, a \in \mathfrak{G}] \end{array}\right\} \text{ for any } a_0 \in \mathfrak{G}.$$

$$\text{(iii)} \qquad W[x_a\,;\, a \in \mathfrak{G}] = [x_{a^{-1}}\,;\, a \in \mathfrak{G}]$$

These operators are unitary and

$$W = W^{-1}, \qquad WU_{a_0}W = V_{a_0}.$$ [48]

In other words, $f \to Wf$ is an involutory spatial automorphism of $\mathfrak{H}$ which interchanges U_{a_0} with V_{a_0}.

PROOF. All assertions are immediately verified.

We introduce, as usual, the complete normalized orthogonal system

$$(5.3.\alpha) \qquad \varphi_a\,, \qquad a \in \mathfrak{G},$$

in $\mathfrak{H}$ where $\varphi_a = [\delta_{a,b}\,;\, b \in \mathfrak{G}]$. Thus $f = [x_a\,;\, a \in \mathfrak{G}]$ is equivalent to $f = \sum_{a\in\mathfrak{G}} x_a\varphi_a$ and

$$(5.3.\beta) \qquad U_{a_0}\varphi_a = \varphi_{aa_0^{-1}}\,, \qquad V_{a_0}\varphi_a = \varphi_{a_0a\cdot}\,, \qquad W\varphi_a = \varphi_{a^{-1}}\,.$$

LEMMA 5.3.2. *Let* I *be the set of all U_{a_0} and* J *the set of all V_{a_0}. Form for each bounded operator A in $\mathfrak{H}$, its matrix $\{\alpha_{a,b}\}(a, b \in \mathfrak{G})$ in the usual way.* [49]

---

[47] This $\mathfrak{H}$ coincides with the $\mathfrak{H}_{\mathfrak{G}}$ of [5], p. 194, Def. 12.1.4. The only difference is that we now write x_a in place of the $f(a)$, loc. cit.

[48] One also verifies immediately

$$U_{a_1}\cdot U_{a_0} = U_{a_1a_0}\,, \qquad V_{a_1}V_{a_0} = V_{a_1a_0}$$

i.e. $a \to U_a$ and $a \to V_a$ are both representations of $\mathfrak{G}$. They correspond in fact to the regular representation of $\mathfrak{G}$.

[49] The definition is

$$({}^{49}, \alpha) \qquad \alpha_{a,b} = (A\varphi_a\,, \varphi_b) = (\varphi_a\,, A^*\varphi_b).$$

Hence

$$\| A\varphi_a \|^2 = \sum_{b\in\mathfrak{G}} | (A\varphi_a\,, \varphi_b) |^2 = \sum_{b\in\mathfrak{G}} | \alpha_{a,b} |^2$$

$$\| A^*\varphi_b \|^2 = \sum_{a\in\mathfrak{G}} | \varphi_a\,, A^*\varphi_b) |^2 = \sum_{a\in\mathfrak{G}} | \alpha_{a,b} |^2$$

i.e.

$$({}^{49}, \beta) \qquad \text{all } \sum_{b\in\mathfrak{G}} | \alpha_{a,b} |^2, \qquad \sum_{a\in\mathfrak{G}} | \alpha_{a,b} |^2 \text{ are finite.}$$

(continued)

Then $A \in \mathbf{I}'$ if and only if $\alpha_{a,b}$ has the form $\alpha_{a,b} = \eta_{ab^{-1}}$ and $A \in \mathbf{J}'$ if and only if $\alpha_{a,b}$ has the form $\alpha_{a,b} = \eta_{a^{-1}b}$ i.e. $\alpha_{a,b}$ depends on the value of ab^{-1} only or on the value of $a^{-1}b$ only, respectively. (We do not determine for which systems η_c, $c \in \mathfrak{G}$ a bounded A to which the given η_c corresponds, actually does exist.)

PROOF. If $A \sim \{\alpha_{a,b}\}$, then (5.3.β) gives $U_{a_0}^{-1}AU_{a_0} \sim \{\alpha_{aa_0^{-1},ba_0^{-1}}\}$.[50] Now $A \in \mathbf{I}'$ means that A commutes with all U_{a_0} i.e. that

$$(5.3.\gamma) \qquad \alpha_{a,b} = \alpha_{aa_0^{-1},ba_0^{-1}} \quad \text{for all} \quad a_0 \in \mathfrak{G}.$$

Write $\eta_c = a_{c,1}$. Then (5.3.γ) with $a_0 = b$ gives

$$(5.3.\delta) \qquad \alpha_{a,b} = \eta_{ab^{-1}}.$$

Conversely (5.3.δ) implies (5.3.γ). Thus (5.3.δ) is characteristic for $A \in \mathbf{I}'$. This proves our statement concerning $\mathbf{I}'$.

If $A \sim \{\alpha_{a,b}\}$ then $WAW \sim \{\alpha_{a^{-1},b^{-1}}\}$.[50] Hence the automorphism $f \to Wf$ of $\mathfrak{H}$ carries (5.3.δ) into

$$(5.3.\epsilon) \qquad \alpha_{a,b} = \eta_{a^{-1}b}.$$

Since it carries $\mathbf{I}$ into $\mathbf{J}$ it takes $\mathbf{I}'$ into $\mathbf{J}'$. Therefore (5.3.ϵ) is characteristic for $A \in \mathbf{J}'$. This proves our statement concerning $\mathbf{J}'$.

LEMMA 5.3.3. $\mathbf{R}(\mathbf{I}) = \mathbf{J}'$, $\mathbf{R}(\mathbf{J}) = \mathbf{I}'$.

PROOF. Clearly V_{a_0} commutes with all U_{a_0}. So $V_{a_0} \in \mathbf{I}'$ and hence $\mathbf{J} \subseteq \mathbf{I}'$. As $\mathbf{I}'$ is a ring, this implies $\mathbf{R}(\mathbf{J}) \subseteq \mathbf{I}'$.

If $A \in \mathbf{I}'$, $B \in \mathbf{J}'$ then $A \sim \{\eta_{ab^{-1}}\}$, $B \sim \{\theta_{a^{-1}b}\}$. Direct computation shows that $AB = BA$ (use the matrix computation rules as mentioned at the end of footnote [49]). Thus $A \in \mathbf{J}''$ and hence $\mathbf{I}' \subseteq \mathbf{J}''$. Since $1 \in \mathbf{J}$, $\mathbf{J}'' = \mathbf{R}(\mathbf{J})$ (cf. [2], p. 397). Therefore $\mathbf{I}' \subseteq \mathbf{R}(\mathbf{J})$.

Finally

$$(Af, \varphi_b) = (f, A^*\varphi_b) = \sum\nolimits_{a \in \mathfrak{G}} (f, \varphi_a)(\varphi_a, A^*\varphi_b)$$
$$= \sum\nolimits_{a \in \mathfrak{G}} \alpha_{a,b}(f, \varphi_a)$$

i.e.

$$({}^{49}, \gamma) \qquad \begin{cases} \text{If} \quad f = \sum_{a \in \mathfrak{G}} x_a \varphi_a, \qquad Af = \sum_{a \in \mathfrak{G}} y_a \varphi_a \\ \text{then } y_b = \sum_{a \in \mathfrak{G}} \alpha_{a,b} x_a. \end{cases}$$

The matrix operations for these (complex numerical) matrices are defined in the same way as for the (operator) matrices of B_p) in §2.4.

[50] Use (^{49}, α) and (5.3. β). Then

$$(U_{a_0}^{-1}AU_{a_0}\varphi_a, \varphi_b) = (AU_{a_0}\varphi_a, U_{a_0}\varphi_b)$$
$$= (A\varphi_{aa_0^{-1}}, \varphi_{ba_0^{-1}})$$
$$(WAW\varphi_a, \varphi_b) = (AW\varphi_a, W\varphi_b) = (A\varphi_{a^{-1}}, \varphi_{b^{-1}})$$

Hence $A \backsim \{\alpha_{a,b}\}$ implies $U_{a_0}^{-1}AU_{a_0} \backsim \{\alpha_{aa_0^{-1},ba_0^{-1}}\}$ and $WAW \backsim \{\alpha_{a^{-1},b^{-1}}\}$.

The results of the two preceding paragraphs imply $\mathbf{R}(\mathbf{J}) = \mathbf{I}'$. The spatial automorphism $f \to Wf$ of $\mathfrak{H}$ interchanges $\mathbf{I}$ with $\mathbf{J}$ and thus we obtain $\mathbf{R}(\mathbf{I}) = \mathbf{J}'$ too. The proof is now complete.

LEMMA 5.3.4. *Put* $\mathbf{M} = \mathbf{R}(\mathbf{I}) = \mathbf{J}'$. *Then* $\mathbf{M}' = \mathbf{R}(\mathbf{J})$. $\mathbf{I}'$ *is a factor if and only if* $\mathfrak{G}$ *fulfills this condition*:

$$\text{(i)} \begin{cases} \textit{For every } a \in \mathfrak{G},\ a \neq 1, \textit{ the class of } a, \\ \qquad\qquad \mathfrak{K}_a = (c^{-1}ac;\ c \in \mathfrak{G}) \\ \textit{is infinite.} \end{cases}$$

PROOF. Clearly $\mathbf{M} = (\mathbf{R}(\mathbf{I}))' = \mathbf{I}'$.

Now consider an $A \in \mathbf{M}\cdot\mathbf{M}'$. Put $A \sim \{\alpha_{a,b}\}$. $A \in \mathbf{M}\cdot\mathbf{M}'$ implies that both

$$(5.3.\zeta) \qquad \alpha_{a,b} = \eta_{a^{-1}b} \quad \text{and} \quad \alpha_{a,b} = \theta_{ab^{-1}}$$

hold. Put $b = 1$. This gives $\eta_{a^{-1}} = \theta_a$. Thus $(5.3.\zeta)$ is equivalent to

$$(5.3.\eta) \qquad \alpha_{a,b} = \eta_{a^{-1}b} = \eta_{ba^{-1}}\,.$$

The second equation in $(5.3.\eta)$ may be rewritten by replacing a, b by b, ab. Then it becomes $\eta_{b^{-1}ab} = \eta_a$ i.e. it states that η_d is a constant for all $d \in \mathfrak{K}_a$. So $(5.3.\eta)$ is equivalent to this

$$(5.3.\theta) \qquad \alpha_{a,b} = \eta_{a^{-1}b}\,,\ \eta_d \text{ is constant for all } d \in \mathfrak{K}_a\,.$$

We now prove that (i) is characteristic.

Sufficiency: Assume that (i) is true. If $a \neq 1$ then $\mathfrak{K}_a$ is infinite. Hence the finiteness of $\sum_{a \in \mathfrak{G}} |\eta_d|^2$ [51] together with $(5.3.\theta)$ imply $\eta_a = 0$ if $a \neq 1$.

$$\text{Thus} \qquad \alpha_{a,b} = \eta_{a^{-1}b} \begin{cases} = \eta_1 & \text{for } a = b \\ = 0 & \text{otherwise.} \end{cases}$$

or $A = \eta\cdot 1$ and so $\mathbf{M}$ is a factor.

Necessity: Assume that (i) is not true. Choose an $a_0 \neq 1$ with $\mathfrak{K}_{a_0}$ finite. Define

$$(5.3.\iota) \qquad \alpha_{a,b} = \eta_{a^{-1}b}\,,\ \eta_d \begin{cases} = 1 & \text{for } a \in \mathfrak{K}_a \\ = 0 & \text{otherwise.} \end{cases}$$

One verifies that the $A \sim \{\alpha_{a,b}\}$ may be expressed as $A = \sum_{d \in \mathfrak{K}_a} U_d$.[52] Thus $A \in \mathbf{M}\cdot\mathbf{M}'$ by $(5.3.\theta)$ and $(5.3.\iota)$.

If $\mathbf{M}$ were a factor, this would necessitate $A = \alpha\cdot 1$ or $A \sim \{\alpha\delta_{a,b}\}$ where $\delta_{a,b} = 1$ for $a = b$ and $\delta = 0$ otherwise. Consequently

$$(5.3.\kappa) \qquad \alpha_{a,b} = \alpha\delta_{a,b}\,.$$

Now $(5.3.\iota)$ and $(5.3.\kappa)$ give for $a = 1$ and $b = a_0$ (remember $a_0 \neq 1$) $\alpha_{1,a_0} = 1$ and $\alpha_{1,a_0} = 0$ respectively. This contradiction shows that $\mathbf{M}$ is not a factor.

---

[51] The finiteness of $\sum_{d \in \mathfrak{G}} |\eta_d|^2$ follows immediately from the first formula of ([49], β) with $a = 1$:

$$\sum_{b \in \mathfrak{G}} |\alpha_{1,b}|^2 = \sum_{b \in \mathfrak{G}} |\eta_b|^2.$$

[52] Observe that the sum can be formed because $\mathfrak{K}_a$ is finite.

This last result forces us to assume that $\mathfrak{G}$ is infinite, since even its classes $\mathfrak{K}_a$, $a \neq 1$ are to be infinite. Groups $\mathfrak{G}$ with this property are easy to construct. We shall give some examples at the end of §5.5 and in §5.6. Clearly no such group can be abelian.

There is, however, another circumstance worth pointing out before we go further. Our construction is obviously nothing but an extension of Frobenius's group numbers to infinite groups.[53] The unitary space $\mathfrak{H}$ and its operators are only technical devices to take care of convergence difficulties. Both the U_a and V_a furnish representations of $\mathfrak{G}$ (cf. footnote [48]), and so both $\mathbf{M} = \mathbf{R}(\mathbf{I})$ and $\mathbf{M}' = \mathbf{R}(\mathbf{J})$ are the equivalents of Frobenius's group numbers. But in the case of finite $\mathfrak{G}$ this procedure does not give a factor. The group numbers notoriously possess other central elements than the $\alpha \cdot 1$. There are as many linearly independent ones as there are different classes $\mathfrak{K}_a$ in $\mathfrak{G}$. The real meaning of our Lemma 5.3.4 appears to be this: The role of all classes in $\mathfrak{G}$, when $\mathfrak{G}$ is finite, is now taken over by the finite classes.[54] For finite $\mathfrak{G}$ this distinction is vacuous, but for the infinite $\mathfrak{G}$ it is now seen to be decisive.

In other words: We have essentially merely extended the construction of Frobenius group numbers to infinite $\mathfrak{G}$. For finite $\mathfrak{G}$ this can never give a factor; indeed, the usefulness of the group numbers lies in that case in the opposite direction. For infinite $\mathfrak{G}$ however, our treatment of the convergence questions makes this construction perfectly adequate to produce factors.

We now determine the case to which these factors $\mathbf{M}$, $\mathbf{M}'$ belong, and a few other characteristics.

The condition (i) in Lemma 5.3.4 is assumed to be fulfilled.

LEMMA 5.3.5. (i) $\mathbf{M}$, $\mathbf{M}'$ *are coupled factors both in case* (II_1).

(ii) *In their standard normalization, the* C *of* [5], *p.* 182, *Theorem* X, *is* 1.

(iii) *If* $A \in \mathbf{M}$, ($A \in \mathbf{M}'$) *then* $Tr_{\mathbf{M}}(A)$, ($Tr_{\mathbf{M}'}(A)$) *has the value* η_1 *for the* η_c *of Lemma* 5.3.2.

PROOF. We proceed in a somewhat changed order.

Define for $A \in \mathbf{M}$, a quantity $T'(A) = \eta_1$ for the η_c of Lemma 5.3.2. Let us check for this $T'(A)$ the properties (i)–(vi′) in [6], p. 218.

(i)–(iii), (v) loc. cit., are obvious.

If $A \sim \{\alpha_{a,b}\} = \{\eta_{a^{-1}b}\}$, $B \sim \{\beta_{a,b}\} = \{\theta_{a^{-1}b}\}$ then clearly $AB \sim \{\gamma_{a,b}\} = \{\zeta_{a^{-1}b}\}$ where $\zeta_a = \sum_{c \in \mathfrak{G}} \eta_c \theta_{ac^{-1}}$. (Use the matrix computation rules indicated at the end of footnote [49].) Hence $T'(AB) = \zeta_1 = \sum_{c \in \mathfrak{G}} \eta_c \theta_{c^{-1}}$. This expression is symmetric in η_a and θ_a. Hence

$$(5.3.\mu) \qquad T'(AB) = T'(BA) = \textstyle\sum_{c \in \mathfrak{G}} \eta_c \theta_{c^{-1}}.$$

This proves (vi), loc. cit., and (vi′) eod., by replacing A, B by $U^{-1}A$, U.

[53] We treat these groups as discrete groups and use everywhere sums not integrals. The familiar generalization of Peter-Weyl to continuous groups and of von Neumann to almost-periodic groups where integrals and means are used, do not show the peculiarities of our present discussion. In those cases the infinite group behaves almost entirely like the finite group.

[54] Concerning the connection between factors and representation theory, cf. also [5], pp. 118–120, Introduction, §3.

As $A^* = \{\overline{\alpha_{b,a}}\} = \{\xi_{a^{-1}b}\}$ with $\xi_a = \overline{\eta_{a^{-1}}}$, (5.3.$\mu$) yields for $B = A^*$

$$T'(A^*A) = \textstyle\sum_{c \in \mathfrak{G}} |\eta_c|^2. \tag{5.3.ν}$$

This proves (iv), (iv′), loc. cit.

Thus all desired relations are established.

If we consider $T'(E)$ only for projections $E \in \mathbf{M}$ then the above relations allow us to assert the validity of (ii), (iii) in [5], p. 165, Def. 8.2.1. Also $T'(1) = 1 \neq 0, \infty$. Hence, by [5], p. 170, Lemma 8.3.5, this $T'(E)$ is a relative dimension function of $\mathbf{M}$ and $\mathbf{M}$ is in a finite case.

The cases (I_n), $n = 1, 2, \cdots$ are excluded, since $\mathbf{M}$ is not a finite ring. Hence $\mathbf{M}$ is in the case (II_1).

Now we conclude by [6], p. 218, Property IV, that $Tr_{\mathbf{M}}(A) = T'(A)$ i.e. that $Tr_{\mathbf{M}}(A) = \eta_1$.

Thus our (i), (iii) hold for $\mathbf{M}$. The spatial automorphism $f \rightarrow Wf$ in $\mathfrak{H}$ shows that they hold for $\mathbf{M}'$ in view of Lemma 5.3.1.

Let us now consider our (ii). Observe that by (5.3.β) $\mathfrak{M}^{\mathbf{M}}_{\varphi_1}$ as well as $\mathfrak{M}^{\mathbf{M}'}_{\varphi_1}$ (cf. [5], p. 143, Def. 5.1.1) contain all φ_a, $a \in \mathfrak{G}$. Hence both are $\mathfrak{H}$. It follows that in the standard normalization.

$$D_{\mathbf{M}}(\mathfrak{M}^{\mathbf{M}'}_{\varphi_1}) = D_{\mathbf{M}'}(\mathfrak{M}^{\mathbf{M}}_{\varphi_1}) = 1.$$

Therefore [5], p. 182, Theorem X, gives $c = 1$ as asserted.

Lemma 5.3.6. *Consider an $A \in \mathbf{M}$ ($A \in \mathbf{M}'$) and its η_c by Lemma 5.3.2. Then we have*

(i) $$[[A]] = (\textstyle\sum_{c \in \mathfrak{G}} |\eta_c|^2)^{\frac{1}{2}}.$$

(ii) $$A = \textstyle\sum_{a \in \mathfrak{G}} \eta_c U_c, \qquad (A = \sum_{a \in \mathfrak{G}} \eta_c V_c)$$

in the sense of metric convergence in $\mathbf{M}$, ($\mathbf{M}'$) irrespective of the order in which the $c \in \mathfrak{G}$ are gone through.

Proof. The spatial automorphism $f \rightarrow Wf$ of $\mathfrak{H}$ interchanges $\mathbf{M}$ and $\mathbf{M}'$. Therefore it suffices to consider the $A \in \mathbf{M}$.

Ad (i): By Equation (5.3.ν) in the proof of Lemma 5.3.5,

$$[[A]]^2 = Tr_{\mathbf{M}}(A^*A) = T'(A^*A) = \textstyle\sum_{c \in \mathfrak{G}} |\eta_c|^2.$$

Hence $[[A]] = (\sum_{c \in \mathfrak{G}} |\eta_c|^2)^{\frac{1}{2}}$.

Ad (ii): Choose any enumeration $a^{(1)}, a^{(2)}, \cdots$ of $\mathfrak{G}$. One verifies immediately that for $A^{(n)} = \sum_{k=1}^{n} \eta_{a(k)} U_{a(k)}$

$$A^{(n)} \sim \{\eta^{(n)}_{a^{-1}b}\} \quad \text{with} \quad \eta^{(n)}_c \begin{cases} = \eta_c & \text{for } c = a^{(1)}, \cdots, a^{(n)} \\ = 0 & \text{otherwise.} \end{cases}$$

Hence

$$A - A^{(n)} \sim \{\eta^{(n)1}_{a^{-1}b}\} \quad \text{with} \quad \eta^{(n)1}_c \begin{cases} = 0 & \text{for } c = a^{(1)}, \cdots, a^{(n)} \\ = \eta_c & \text{otherwise.} \end{cases}$$

Consequently (i) above gives

$$[[A - A^{(n)}]] = (\textstyle\sum_{c \in \mathfrak{G}} | \eta_c^{(n)1} |^2)^{\frac{1}{2}} = (\sum_{c \in \mathfrak{G},\, c \neq a^{(1)}, \cdots, a^{(n)}} | \eta_c |^2)^{\frac{1}{2}}$$
$$= (\textstyle\sum_{i=n+1}^{\infty} | \eta_{a^{(i)}} |^2)^{\frac{1}{2}}$$

This is equivalent to

$$(5.3.o) \qquad [[A - \textstyle\sum_{k=1}^{n} \eta_{a^{(k)}} U_{a^{(k)}}]] = (\sum_{k=n+1}^{\infty} | \eta_{a^{(k)}} |^2)^{\frac{1}{2}}.$$

Now $\sum_{k=1}^{\infty} | \eta_{a^{(k)}} |^2 = \sum_{c \in \mathfrak{G}} | \eta_c |^2$ is finite by (i) above. Hence the right hand side of (5.3.o) tends to 0 as $n \to \infty$. This means that $\sum_{k=1}^{\infty} \eta_{a^{(k)}} U_{a^{(k)}}$ converges metrically to A.

Since this holds for any enumeration $a^{(1)}, a^{(2)}, \cdots$ of $\mathfrak{G}$, our assertion is established.

§5.4 As we pointed out, the construction of §5.3 is closely related to that of [5], pp. 192–209, Part IV. The group $\mathfrak{G}$ plays the same role in both cases, and the difference consists in the disappearance of the space S.

This space played an important part in the construction referred to above: Every $a \in \mathfrak{G}$ had to induce a mapping $x \to xa$ in S (cf. loc. cit., p. 195, Def. 12.1.5). The absence of S from our present construction may also be interpreted by saying that S is now a one-element set: $S = (x_0)$ and that for every $a \in \mathfrak{G}$, $x_0 a = x_0$.

Now the difference between our two constructions can be expressed as follows: In [5], $\mathfrak{G}$ was entirely unrestricted, but S had to obey (among other things) (iii) in Def. 12.1.5, loc. cit., p. 195: for $a \neq 1$ and $x \in S$ always $xa \neq x$.[55] In §5.3, $\mathfrak{G}$ had to fulfill (i) of Lemma 5.3.4, but the (iii) referred to was most flagrantly violated.[56, 57] Thus our two procedures correspond to two different ways of securing the factor character of **M** in what is fundamentally the same construction. In one case the whole burden of the necessary restrictions was thrown on S in the other case, on $\mathfrak{G}$.

It would be possible to use a more general arrangement of which both these procedures are special cases. The necessary restriction would then be of a mixed type, affecting the structure of S and $\mathfrak{G}$.

We shall not consider this question in more detail at this time.

§5.5 A second connection between our two constructions which deserves some attention can be obtained as follows.

Let an arbitrary countably infinite group $\mathfrak{G}$ be given. We shall try to use it

---

[55] Except for an x set of measure zero.

[56] We have $x_0 a = x_0$ and $\mu((x_0)) = \mu(S) \neq 0$.

[57] There is no difference between the two constructions as regards the other conditions of Def. 12.1.5, loc. cit., p. 195. In particular, $\mathfrak{G}$ is ergodic in our present construction, since $S = (x_0)$ is a one-element set.

for the construction of [5] referred to above, that is to find an S which fulfills all preliminary conditions of that construction in conjunction with the given $\mathfrak{G}$.

This construction runs as follows:[58]

(I) Let S be the set of all systems $x = [\alpha_a ; a \in \mathfrak{G}]$ where each $\alpha_a = 0, 1$. Let $\mathfrak{S}$ be the set of those $x = [\alpha_a ; a \in \mathfrak{G}]$ for which $\alpha_a = 0$ except for a finite number of a's.

Define in S: If $x = [\alpha_a ; a \in \mathfrak{G}]$, $y = [\beta_a ; a \in \mathfrak{G}]$ then $x \oplus y = [\gamma_a ; a \in \mathfrak{G}]$ where $\gamma_a = \alpha_a + \beta_a (\text{mod } 2)$, $\gamma_a = 0$ or 1. S is clearly a (commutative) group with the "unit" $0 = [0; a \in \mathfrak{G}]$ and $\mathfrak{S}$ is an enumerably infinite subgroup of S.

Define further in S: If $x = [\alpha_a ; a \in \mathfrak{G}]$ and $a_0 \in \mathfrak{G}$ then $xa_0 = [\alpha_{a_0 a} ; a \in \mathfrak{G}]$. Then the mappings $x \to xa_0$ represent the group $\mathfrak{G}$ by permutations of S, all of which carry $\mathfrak{S}$ into itself.

Now choose an arbitrary but fixed enumeration $a^{(1)}, a^{(2)}, \cdots$ of $\mathfrak{G}$. For S we use the mapping

$$\Xi \qquad x = [\alpha_a; a \in \mathfrak{G}] \to \xi(x) = \sum_{m=1}^{\infty} \frac{\alpha_{a^{(m)}}}{2^m}$$

of S on the numerical interval $0 \leqq \xi \leqq 1$. Except for the Ξ image of $\mathfrak{S}$ the set of all dyadically rational numbers, which is a set of Lebesgue measure 0 this mapping is one-to-one. So the common (exterior) Lebesgue measure in $0 \leqq \xi \leqq 1$ is mapped by the inverse of Ξ on a Lebesgue measure in S with all its usual properties.

We shall now consider $\mathfrak{G}$ as the "group" and S (with the "mappings" $x \to xa_0$ for $x \in S$, $a_0 \in \mathfrak{G}$) as the "space" described in [5], pp. 192–195. In the sense of [5], p. 195, Def. 12.1.5, $\mathfrak{G}$ is an m-group and ergodic in S.

First we show the m-group character. Ad (i), loc. cit. This is concerned with the preservation of outer measure of the measure determining sets $\{T_i\}$. In view of the definition of measure in S given in the previous paragraph, we may take for a set of T_i's the set of images, under the inverse of Ξ, of ξ-intervals $k/2^m \leqq \xi < (k+1)/2^m$. An image of a ξ-interval $k/2^m \leqq \xi < (k+1)/2^m$ is given by specifying the values of $\alpha_{a^{(1)}}, \ldots, \alpha_{a^{(m)}}$. Hence it is mapped by $x \to xa_0$ onto a set, in which the values of $\alpha_{a_0^{-1}a^{(1)}}, \ldots, \alpha_{a_0^{-1}a^{(m)}}$ are specified. Choose n so that $a_0^{-1}a^{(1)}, \cdots, a_0^{-1}a^{(m)}$ occur among the $a^{(1)}, \cdots, a^{(n)}$. Then the above set is the sum of 2^{n-m} sets, each of which is defined by specifying the values of $\alpha_{a^{(1)}}, \ldots, \alpha_{a^{(n)}}$, i.e. each of which corresponds to an interval $l/2^n \leqq \xi < (l+1)/2^n$. Thus the set corresponding to the ξ interval $k/2^m \leqq \xi < (k+1)/2^m$ having measure $1/2^m$ is mapped on a set of measure $2^{n-m}/2^n = 2^{-m}$. Thus the measure of each such is conserved. Ad (ii) which states that $(xa_0)b_0 = x(a_0b_0)$ can be verified by a direct computation.[59] Ad (iii), states that if $a_0 \neq 1$, $xa_0 = x$ holds

[58] The construction (I)–(V) which follows is in many respects analogous to the construction (I)–(VII) in [7], pp. 72–77. The main differences are these: We use $\mathfrak{G}$ and not $\mathfrak{S}$ as the group of mappings in S, and the latter parts of the two constructions pursue different aims.

[59] Let $x = [\alpha_a, a \in \mathfrak{G}]$, $xa_0 = [\beta_a, a \in \mathfrak{G}]$, $(xa_0)b_0 = [\gamma_a, a \in \mathfrak{G}]$. Then $\beta_a = \alpha_{a_0a}$ and $\gamma_a = \beta_{b_0a} = \alpha_{a_0(b_0a)} = \alpha_{(a_0b_0)a}$. Hence $(xa_0)b_0 = [\alpha_{(a_0b_0)a}, a \in \mathfrak{G}] = x(a_0b_0)$.

only for a set of measure zero. For $x = [\alpha_a; a \in \mathfrak{G}]$, $xa_0 = x$ is equivalent to $\alpha_{a_0a} = \alpha_a$ for all a. But since $a_0 \in \mathfrak{G}$, $a_0 \neq 1$ we can construct an infinite sequence $a_1, a_2, \cdots$ such that the $a_1, a_2, \cdots, a_0a_1, a_0a_2, \cdots$ are all different.[60] For any m, let $n = n(m)$ be such that $a^{(1)}, \cdots, a^{(n)}$ includes $a_1, \cdots, a_m, a_0a_1, \cdots, a_0a_m$. Then the x-set $S^{(m)}$ for which $\alpha_{a_0a_i} = \alpha_{a_i}$ for $i = 1, \cdots, m$ consists of 2^{n-m} sets $I^{(n)}$ in each of which the $\alpha_{a^{(1)}}, \ldots, \alpha_{a^{(n)}}$ have specified values. Such a set $I^{(n)}$ is the image of a ξ interval of measure 2^{-n} and hence $S^{(m)}$ has measure $2^{n-m} \cdot 2^{-n} = 2^{-m}$. Now the x-set for which $\alpha_{a_0a} = \alpha_a$ for all $a \in \mathfrak{G}$ is in $S^{(m)}$ for every m. Thus it has measure zero. Since this is also the set for which $xa_0 = x$ we have shown, Ad (iii).

The ergodicity will be established in (III) below.

(II) Form for these S, $\mathfrak{G}$ the spaces $\mathfrak{H}_S$ and $\mathfrak{H}_{\mathfrak{G}S}$ of all complex-valued functions $f(x)$ resp. $F(x, a)$, $x \in S$, $a \in \mathfrak{G}$ which are measurable in x and with a finite $\int_S |f(x)|^2 dx$ resp. $\sum_{a \in \mathfrak{G}} \int_S |F(x, a)|^2 dx$. Thus $(f, g) = \int_S f(x)\overline{g(x)}\, dx$, $(F, G) = \sum_{a \in \mathfrak{G}} \int_S F(x, a)\overline{G(x, a)}\, dx$ (cf. [5], p. 194). Form the bounded operators

$$\left.\begin{aligned} U_{a_0} f(x) &= f(xa_0) \\ \overline{U_{a_0}} F(x, a) &= F(xa_0, aa_0) \end{aligned}\right\} a_0 \in \mathfrak{G}.$$

$$\overline{L_{\varphi(x)}} F(x, a) = \varphi(x) F(x, a) \quad (\varphi(x) \text{ bounded and measurable})$$

(loc. cit., pp. 196, 198–199) and the ring $\mathbf{M}$ in $\mathfrak{H}_{\mathfrak{G}S}$ which the $\overline{U_{a_0}}$, $\overline{L_{\varphi(x)}}$ generate (loc. cit., p. 200):

$$(5.5.\alpha) \qquad \mathbf{M} = \mathbf{R}(\overline{U_{a_0}}, \overline{L_{\varphi(x)}}, a_0 \in \mathfrak{G}, \varphi(x) \text{ bounded and measurable}).$$

Before we go on we form the following system of functions in $\mathfrak{H}_{\mathfrak{G}S}$

$$w_{\bar{x}}(x) = (-1)^{\sum_{a \in \mathfrak{G}} \bar{\alpha}_a \alpha_a} \qquad (\bar{x} = [\bar{\alpha}_a, a \in \mathfrak{G}] \in \mathfrak{S}, x = [\alpha_a, a \in \mathfrak{G}] \in S.)^{62}$$

It is easy to verify that these functions (for all $\bar{x} \in \mathfrak{S}$) form a complete normalized orthogonal system in $\mathfrak{H}_S$.[63] Consequently the functions

$$F_{\bar{x}, \bar{a}}(x, a) = \omega_{\bar{x}}(x) \delta_{a, \bar{a}}, \qquad (x \in \mathfrak{S}, \bar{a} \in \mathfrak{G})$$

form a complete normalized orthogonal system in $\mathfrak{H}_{\mathfrak{G},S}$.

---

[60] If the $a_0, a_0^2, \cdots$ are all distinct, let $a_i = a_0^{2i-1}$. The result readily follows in this case. Otherwise, a_0 is of finite order, $p = 2, 3, \cdots$ ($p \neq 1$ since $a_0 \neq 1$). For each a we can define S_a as the set of $b \in \mathfrak{G}$ for which $b = a_0^k a$ for some $k = 0, 1, \cdots, p - 1$. It is easily seen how we can choose a sequence $a_1, a_2, \cdots$ such that the S_{a_i} are mutually exclusive. For this sequence we have that the $a_1, a_2, \cdots, a_0a_1, a_0a_2, \cdots$ are all different. This readily follows from the facts that the S_{a_i} are mutually exclusive and $a_0 \neq 1$.

[62] Owing to $\bar{x} = [\bar{\alpha}_a, a \in \mathfrak{G}] \in \mathfrak{S}$ we have $\bar{\alpha}_a \neq 0$ only for a finite set of a's. Hence the sum $\sum_{a \in \mathfrak{G}} \bar{\alpha}_a \alpha_a$ is really finite.

[63] In this particular case the group nature of $\mathfrak{G}$ is irrelevant. Hence we can replace $\mathfrak{G}$ by any other countably infinite set. We replace the $a \in \mathfrak{G}$ by $a = 1, 2, \cdots$. After this substitution our $\omega_{\bar{x}}(x)$ become the well-known complete normalized orthogonal set of Walsh-Rademacher functions. It was also considered in [7], p. 74, under the name of $\omega_a(x)$.

A familiar argument shows that the $\varphi(x)$ in (5.5.α) can be restricted considerably without changing the ring $\mathbf{M}$. Indeed:

First it suffices, for reasons of continuity, to consider those $\varphi(x)$ only which are finite linear aggregates of the $\omega_{\bar{x}}(x)$.[64] Second, we may now restrict the $\varphi(x)$ to the $\omega_{\bar{x}}(x)$ themselves. So we have

$$(5.5.\beta)\qquad \mathbf{M} = \mathbf{R}(\bar{U}_{a_0}, \overline{L_{\omega_{\bar{x}}(x)}};\ a_0 \in \mathfrak{G},\ \bar{x} \in \mathfrak{S}).$$

And as 1 occurs among the $\bar{U}_{a_0}$ (for $a_0 = 1$) and among the $\overline{L_{\omega_{\bar{x}}(x)}}$ (for $\bar{x} = 0$), we can equally write:

$$(5.5.\gamma)\qquad \mathbf{M} = \mathbf{R}(\bar{U}_{a_0}\overline{L_{\omega_{\bar{x}}(x)}};\ a_0 \in \mathfrak{G},\ \bar{x} \in \mathfrak{S}).$$

(III) Clearly

$$U_{a_0}\omega_{\bar{x}}(x) = \omega_{\bar{x}}(xa_0) = \omega_{\bar{x}a_0^{-1}}(x).$$

Now consider any $f \in \mathfrak{H}_S$ with $U_{a_0}f = f$ for all $a_0 \in \mathfrak{G}$. Use the orthogonal expansion $f(x) = \sum_{\bar{x}\in\mathfrak{S}} \lambda_{\bar{x}}\omega_{\bar{x}}(x)$. Then the above formula gives $f = \sum_{\bar{x}\in\mathfrak{S}} \lambda_{\bar{x}}\omega_{\bar{x}}$, $U_{a_0}f = \sum_{\bar{x}\in\mathfrak{S}} \lambda_{\bar{x}}\omega_{\bar{x}a_0^{-1}} = \sum_{\bar{x}\in\mathfrak{S}} \lambda_{\bar{x}a_0}\omega_{\bar{x}}$. Hence $\lambda_{\bar{x}} = \lambda_{\bar{x}a_0}$. As $\|f\|^2 = \sum_{\bar{x}\in\mathfrak{S}} |\lambda_{\bar{x}}|^2$ is finite, this implies that $\lambda_{\bar{x}} = 0$ if there are infinitely many distinct $\bar{x}a_0$, $a_0 \in \mathfrak{G}$. Now this is the case for all $\bar{x} \in \mathfrak{S}$, $\bar{x} \neq 0$.[65] So $\bar{x} \neq 0$ implies $\lambda_{\bar{x}} = 0$. Thus $f(x) = \lambda_0\omega_0(x) = \lambda_0$.

So $U_{a_0}f = f$ (for all $a_0 \in \mathfrak{G}$) implies that f is a constant. Hence $\mathfrak{G}$ is ergodic.

(IV) Clearly

$$\omega_{\bar{x}}(x)\omega_{\bar{y}}(x) = \omega_{\bar{x}\oplus\bar{y}}(x)$$

$$\overline{L_{\omega_{\bar{x}}(x)}}F_{\bar{y},\bar{b}}(x, a) = \omega_{\bar{x}}(x)F_{\bar{y},\bar{b}}(x, a) = F_{\bar{x}\oplus\bar{y},\bar{b}}(x, a).$$

Next

$$\bar{U}_{\bar{a}}F_{\bar{y},\bar{b}}(x, a) = F_{\bar{y},\bar{b}}(x\bar{a}, a\bar{a}) = F_{\bar{y}\bar{a}^{-1},\bar{b}\bar{a}^{-1}}(x, a).$$

Combining these gives

$$(5.5.\epsilon)\qquad \bar{U}_{\bar{a}}\bar{L}_{\omega_{\bar{x}}(x)}F_{\bar{y},\bar{b}}(x, a) = F_{(\bar{x}\oplus\bar{y})\bar{a}^{-1},\bar{b}\bar{a}^{-1}}(x, a).$$

Now treat the index $\bar{y}$, $\bar{b}$ of $F_{\bar{y},\bar{b}}$ as one pair. Make these pairs into a group $\{\mathfrak{S}, \mathfrak{G}\}$ by defining

$$(5.5.\zeta)\qquad \{\bar{y}, \bar{b}\}\{\bar{z}, \bar{c}\} = \{\bar{y}\bar{c} \oplus \bar{z}, \bar{b}\bar{c}\}.$$

---

[64] Cf. the argument of [7], pp. 73–74 (II).

[65] We prove this as follows: Put $\bar{x} = [\bar{x}_a, a \in \mathfrak{G}]$. Since $\bar{x} \neq 0$ there is an a^* such that $\bar{x}_{a^*} \neq 0$. Now as in footnote[60], we can construct an infinite sequence of the a's $\in \mathfrak{G}$ such that a_1a^*, a_2a^* are all different. Then $\bar{x}a_i^{-1} = [\bar{x}_{a_i^{-1}a}, a \in \mathfrak{G}]$ has for its a_ia^* component $\bar{x}_{a^*} = 1$. Thus if we consider the $\bar{x}a_1^{-1}$, $\bar{x}a_2^{-1}$, $\cdots$ we find that there is an infinite number of a's $\in \mathfrak{G}$ such that the a component is not zero for at least one of the $\bar{x}a_i^{-1}$. Since each $\bar{x}a_i^{-1}$ can have at most a finite number of components different from zero, there must be an infinite number of distinct $\bar{x}a_i^{-1}$.

As $\bar{x} \oplus \bar{x} = 0$ (5.5.ζ) implies that

$$\{\bar{y}, \bar{b}\}\{\bar{x}, \bar{a}\}^{-1} = \{(\bar{x} \oplus \bar{y})\bar{a}^{-1}, \bar{b}\bar{a}^{-1}\}.$$

Hence (5.5.ϵ) becomes

$$(5.5\eta) \qquad U_{\bar{a}} L_{\omega_{\bar{x}}(x)} F_{\{\bar{y},\bar{b}\}}(x, a) = F_{\{\bar{y},\bar{b}\}\{\bar{x},\bar{a}\}^{-1}}(x, a).$$

(V) Apply the construction of §5.3 to the group $\{\mathfrak{S}, \mathfrak{G}\}$ of (5.5.ζ) above, instead of its $\mathfrak{G}$. This is possible since $\{\mathfrak{S}, \mathfrak{G}\}$ fulfills (i) in Lemma 5.3.4; i.e. if $\{\bar{x}, \bar{a}\} \neq \{0, 1\}$ then there exist infinitely many different $\{\bar{y}, \bar{b}\}^{-1}\{\bar{x}, \bar{a}\}\{\bar{y}, \bar{b}\}$. Indeed, if $\bar{x} \neq 0$ then form $\{0, \bar{b}\}^{-1}\{\bar{x}, \bar{a}\}\{0, \bar{b}\} = \{\bar{x}\bar{b}, \bar{b}^{-1}\bar{a}\bar{b}\}$. There are infinitely many distinct $\bar{x}\bar{b}$ for $\bar{b} \in \mathfrak{G}$. Cf. footnote [65]. Thus our statement is true for $\bar{x} \neq 0$. If however $\bar{x} = 0$, then necessarily $\bar{a} \neq 1$. Form $\{\bar{y}, 1\}\{0, \bar{a}\}\{\bar{y}, 1\} = \{\bar{y}\bar{a} \oplus \bar{y}, \bar{a}\}$. There exist infinitely many different $\bar{y}\bar{a} \oplus \bar{y}$.[66]

Now write $\mathfrak{H}_I$, $\mathbf{M}_I$ for its $\mathfrak{H}$, $\mathbf{M}$. We have

$$(5.5.\theta) \qquad \bar{\mathbf{M}}_I = \mathbf{R}(U_{\{\bar{x},\bar{a}\}}\,;\, \bar{x} \in \mathfrak{S}, \bar{a} \in \mathfrak{G})$$

and by (5.3.β)

$$(5.5.\iota) \qquad U_{\{\bar{x},\bar{a}\}}\varphi_{\{\bar{y},\bar{b}\}} = \varphi_{\{\bar{y},\bar{b}\}\{\bar{x},\bar{a}\}^{-1}}.$$

We can map $\mathfrak{H}_{\mathfrak{G}S}$ isomorphically on $\mathfrak{H}_I$ by carrying the complete orthonormalized system $F_{\{\bar{y},\bar{b}\}}$ of the former into the set of $\varphi_{\{\bar{y},\bar{b}\}}$ of the latter. Then comparison of (5.5.ι) and (5.5.η) shows that this isomorphism carries $U_{\bar{a}}L_{\omega_{\bar{x}}(x)}$ into $U_{\{\bar{x},\bar{a}\}}$. Hence (5.5.γ) and (5.5.θ) show that it carries $\mathbf{M}$ into $\mathbf{M}_I$. Thus $\mathbf{M}$ and $\mathbf{M}_I$ are spatially isomorphic.

Summing up:

LEMMA 5.5.1. *Given any countably infinite group* $\mathfrak{G}$ *define* S, $\mathfrak{S}$ *as described in* (I) *above, and the group* $\mathfrak{G}_I = \{\mathfrak{S}, \mathfrak{G}\}$ *as described by* (5.5.ζ) (IV) *above. Then the construction of* [5], *pp.* 192–209, *Part IV, can be applied to* $\mathfrak{G}$ *and* S *and our construction of* §5.3 *to* $\{\mathfrak{S}, \mathfrak{G}\}$ (*in place of its* $\mathfrak{G}$).

These two constructions produce spatially isomorphic factors $\mathbf{M}$.

The mappings

$$(5.5.\kappa) \qquad x\{\bar{y}, \bar{b}\} = x\bar{b} + \bar{y}$$

provide a convenient representation of the group $\{\mathfrak{S}, \mathfrak{G}\}$ of (5.5.ζ) in (V). If we specify $x \in S$ then (5.5.κ) appears as a permutation of S. But as it maps $\mathfrak{S}$ on itself, we can also prescribe $x \in \mathfrak{S}$ and view (5.5.κ) as a permutation of $\mathfrak{S}$.

(5.5.κ) makes it clear that $\{\mathfrak{S}, \mathfrak{G}\}$ is the simplest combination of $\mathfrak{S}$ and $\mathfrak{G}$ apart from the "direct group product".

Our above result can also be interpreted in this way:

We start with an arbitrary (countably infinite) group $\mathfrak{G}$ and should like to apply the construction of §5.3. Since $\mathfrak{G}$ may not fulfill the condition (i) of

---

[66] Put $\bar{y}_b = [\delta_{a,b}, a \in \mathfrak{G}], \in \mathfrak{S}$. As $\bar{a} \neq 1$, $\bar{y}_b\bar{a} \oplus \bar{y}_b = \bar{y}_c\bar{a} \oplus \bar{y}_c$ means either $b = c$ or $\bar{a}b = c$, $\bar{a}c = b$. In other words: If b runs over all $\mathfrak{G}$ there can never coincide more than two $\bar{y}_b a \oplus \bar{y}_b$. Now $\mathfrak{G}$ is infinite. Hence there are infinitely many different $\bar{y}_b\bar{a} \oplus \bar{y}_b$.

Lemma 5.3.4 this cannot always be done directly. We therefore propose to expand $\mathfrak{G}$ to a group $\mathfrak{G}_I$ which fulfills that condition. (5.5.ζ) provides a very simple generally valid mechanism to do just that. This has the further interesting consequence that within **M**, the set of group numbers for any $\mathfrak{G}$, the metric closure of a subalgebra is equivalent to the topological closure for any of the ring topologies. For we extend **M** to $\mathbf{M}_I$ as above, and Theorem I (of §1.6 above) shows that our statement is true in the extended space $\mathfrak{H}_I$. Since any subalgebra of **M** will be reduced by the original space $\mathfrak{H}$, topological closure in $\mathfrak{H}_I$ will yield topological closure. Similarly the discussion of Chapter I may be carried through on the assumption that **M** is a group algebra.

§5.6 The analogue of Lemma 5.2.2 is true for the construction of §5.3 too, and this time it is even easier to prove.

LEMMA 5.6.1. *The ring* **M** *of* §5.3 *is approximately finite if* $\mathfrak{G}$ *possesses the property* (*i*) *of Lemma* 5.2.2 (*besides the property* (*i*) *of Lemma* 5.3.4).

PROOF. Form the $\mathfrak{G}_1, \mathfrak{G}_2, \cdots$ of (i) in Lemma 5.2.2. Now

$$\mathbf{M} = \mathbf{R}(U_a\,;\, a \in \mathfrak{G}).$$

Hence

$$\mathbf{M} = \mathbf{R}(\mathbf{N}_n\,;\, n = 1, 2, \cdots)$$

with

$$\mathbf{N}_n = \mathbf{R}(U_a\,;\, a \in \mathfrak{G}_n).$$

Thus $\mathbf{N}_1 \subseteq \mathbf{N}_2 \subseteq \cdots$ and $\mathbf{N}_n$ is of finite order because $\mathfrak{G}_n$ is finite. Therefore **M** is approximately finite by Def. 4.6.1.[46]

Observe that if $\mathfrak{G}$ fulfills (i) in Lemma 5.2.2 (but not necessarily (i) in Lemma 5.3.4) then the group $\mathfrak{G}_I = \{\mathfrak{S}, \mathfrak{G}\}$ of Lemma 5.5.1 possesses that property too. Put $\mathfrak{G} = \mathfrak{G}_1 + \mathfrak{G}_2 + \cdots$, $\mathfrak{G}_1 \subseteq \mathfrak{G}_2 \subseteq \cdots$ all $\mathfrak{G}_n$ finite. Let $\mathfrak{S}_n$ be the set of all $x = [\bar{\alpha}_a, a \in \mathfrak{G}]$ with $\bar{\alpha}_a = 0$, when a is not in $\mathfrak{G}_n$. Then $\mathfrak{S}_1 \subseteq \mathfrak{S}_2 \subseteq \cdots$ and all the $\mathfrak{S}_n$ are finite. So $\{\mathfrak{S}, \mathfrak{G}\} = \{\mathfrak{S}_1, \mathfrak{G}_1\} + \{\mathfrak{S}_2, \mathfrak{G}_2\} + \cdots$, $\{\mathfrak{S}_1, \mathfrak{G}_1\} \subseteq \{\mathfrak{S}_2, \mathfrak{G}_2\} \subseteq \cdots$ and all the $\{\mathfrak{S}_n, \mathfrak{G}_n\}$ are finite.

Thus it is easy to form groups $\mathfrak{G}$ which fulfill both (i) in Lemma 5.2.2 and (i) in Lemma 5.3.4. The following example is more direct.

LEMMA 5.6.2. *The group* $\tilde{\mathfrak{G}}$ *of all those permutations of* $(1, 2, \cdots)$ *which move only a finite number of elements, fulfills both* (*i*) *in Lemma* 5.2.2 *and* (*i*) *in Lemma* 5.3.4.

PROOF. Ad (i) in Lemma 5.3.4.: Given an $a \in \tilde{\mathfrak{G}}$, $a \neq 1$ choose n so that a moves no number other than $1, \cdots, n$. Assume that a does move i (which is therefore $= 1, \cdots, n$). Let c_j be the transposition of i and j where $j = n + 1, n + 2, \cdots$. Then the only $k = n + 1, n + 2, \cdots$ moved by $c_j^{-1} ac_j$ is $k = j$. Thus the $c_j^{-1}ac_j$, $j = n + 1, n + 2, \cdots$ are pairwise different and so $\mathfrak{K}_a$ is infinite.

Ad (i) in Lemma 5.2.2.: Let $\tilde{\mathfrak{G}}_n$ be the set of those permutations which move no number other than $1, \cdots, n$. Then $\tilde{\mathfrak{G}} = \tilde{\mathfrak{G}}_1 + \tilde{\mathfrak{G}}_2 + \cdots$, $\tilde{\mathfrak{G}}_1 \subseteq \tilde{\mathfrak{G}}_2 \subseteq \cdots$ and each $\tilde{\mathfrak{G}}_n$ is a finite subset of $\mathfrak{G}$.

The analysis of the invariants (1), (2) of §4.8, gives no new information. As to (1) all examples of [5] as well as those of §5.3 have $\bar{\mathbf{M}}^c = \bar{\mathbf{M}}$. Indeed: Neither of these definitions ever mentions any specific non-real (complex) number. Hence $\mathfrak{H}_c$, $\mathbf{M}_c$ coincide with $\mathfrak{H}$, $\mathbf{M}$ so that $\bar{\mathbf{M}}^c = \bar{\mathbf{M}}^{`} = \bar{\mathbf{M}}$ (cf. §2.3).

Thus we possess no examples of a factor $\mathbf{M}$ in the case (II_1) with $\bar{\mathbf{M}}^c \neq \bar{\mathbf{M}}$. Possibly a generalization of the construction of 5.3, with the introduction of complex numerical factors in (i), (ii) of Lemma 5.3.1 (and (5.3.β) after that lemma) might achieve this.[67] But we have not obtained so far any conclusive results in this respect.

As to (2), $\mathfrak{G}(\bar{\mathbf{M}})$ we know nothing beyond Theorem XV.

Chapter VI. Non-Approximately Finite Factors

§6.1 As pointed out at the end of §4.8 and again at the end of §5.6, we have not succeeded so far to establish the not approximately finite character of any factor with the help of the invariants (1), (2) of §4.8. We shall now prove the existence of not approximately finite factors, but it is necessary to introduce a new invariant in order to achieve this aim.

Throughout what follows, $\mathbf{M}$ is again a factor in case (II_1).

Definition 6.1.1. $\mathbf{M}$ *has the property* Γ *if this is true:*

Given any system, $A_1, \cdots, A_m \in \mathbf{M}$ *and any* $\epsilon > 0$ *there exists a unitary* $U = U(A_1, \cdots, A_m, \epsilon) \in \mathbf{M}$ *with*

$$Tr_{\mathbf{M}}(U) = 0$$

and

$$[[U^{-1}A_kU - A_k]] < \epsilon \quad for \quad h = 1, \cdots, m.$$ [68, 69]

This property Γ is clearly a purely algebraic property—i.e. one concerning the algebraic type $\bar{\mathbf{M}}$ only.

---

[67] Corresponding to the technique of "multipliers" in forming skew fields over a given center in abstract algebra.

[68] The requirement $Tr_{\mathbf{M}}(U) = 0$ is necessary, since otherwise we could choose $U = 1$. Outright commutativity of U with $A_1, \cdots, A_m$ (instead of the last inequality) would not do for the following reason: If $\mathbf{M} = \mathbf{R}(A_1, \cdots, A_m)$ then this commutativity would mean $U \in \mathbf{M}'$. Hence $U \in \mathbf{M}\cdot\mathbf{M}'$, $U = \alpha\cdot 1$. Since $Tr_{\mathbf{M}}(U) = 0$ this would mean $U = 0$, contradicting the unitarity of U. Hence no $\mathbf{M} = \mathbf{R}(A_1, \cdots, A_m)$ could then have this property. Now it can be shown that an approximately finite $\mathbf{M} = \mathbf{R}(A_1, A_2)$ for two suitable A_1, A_2. Hence an approximately finite $\mathbf{M}$ would not have this property—but we need Lemma 6.1.2.

[69] This property could be further subdivided, according to the values of m and of expressions like $\epsilon/\mathrm{Max}\,(|||\,A_1\,|||, \cdots, |||\,A_m\,|||)$. We could also require $A_1, \cdots, A_m$ to be unitary, etc. For our present purpose, however, these refinements are not needed.

Observe first this

LEMMA 6.1.1. *Let a group $\mathfrak{G}$ fulfilling (i) in Lemma 5.3.4 be given, which possesses this property:*

$$\text{(i)} \quad \begin{cases} \textit{Given any system } a^{(1)}, \cdots, a^{(n)} \in \mathfrak{G} \textit{ there exists} \\ a\ c_0 \in \mathfrak{G},\ c_0 \neq 1 \textit{ which commutes with } a^{(1)}, \cdots, a^{(n)}. \end{cases}$$

Then the **M** *of* §5.3 *possesses the property* Γ.

PROOF. Let $A_1, \cdots, A_m \in \mathbf{M}$ and $\epsilon > 0$ be given as indicated in Def. 6.1.1. Following (ii) in Lemma 5.3.6, we can choose $a^{(1)}, \cdots, a^{(n)} \in \mathfrak{G}$ and $\zeta_{h,1}, \cdots, \zeta_{h,n}$ (these are the $\eta_{a^{(1)}}, \cdots, \eta_{a^{(n)}}$ mentioned there for $A = A_h$ for $h = 1, \cdots, m$) so that

$$(6.1.\alpha) \qquad [[A_h - \textstyle\sum_{i=1}^{n} \zeta_{h,i} U_{a^{(i)}}]] < \epsilon/2 \qquad \text{for } h = 1, \cdots, m.$$

Now apply our assumption (i) to these $a^{(1)}, \cdots, a^{(n)}$ obtaining a $c_0 \in \mathfrak{G}$, $c_0 \neq 1$ which commutes with them.

Then U_{c_0} and $U_{a^{(i)}}$ commute (cf. footnote [48]) and so

$$U_{c_0}^{-1}(A_h - \textstyle\sum_{i=1}^{n} \zeta_{h,i} U_{a^{(i)}})U_{c_0} = U_{c_0}^{-1} A U_{c_0} - \textstyle\sum_{i=1}^{n} \zeta_{h,i} U_{a^{(i)}}.$$

Hence (6.1.α) yields

$$(6.1.\beta) \qquad [[U_{c_0}^{-1} A U_{c_0} - \textstyle\sum_{i=1}^{n} \zeta_{h,i} U_{a^{(i)}}]] < \epsilon/2$$

and (6.1.α) and (6.1.β) yield together

$$(6.1.\gamma) \qquad [[U_{c_0}^{-1} A_h U_{c_0} - A_h]] < \epsilon.$$

$$\text{Clearly } U_{c_0} \sim \{\eta'_{ab^{-1}}\} \quad \text{where } \eta'_c \begin{cases} = 1 & \text{for } c = c_0 \\ = 0 & \text{otherwise.} \end{cases}$$

As $c_0 \neq 1$ this implies $\eta'_1 = 0$ i.e.

$$(6.1.\delta) \qquad Tr_{\mathbf{M}}(U_{c_0}) = 0.$$

Thus (6.1.γ), (6.1.δ) show that $U = U_{c_0}$ meets all our requirements.

We can now prove

LEMMA 6.1.2. *Approximate finiteness implies the Property* Γ.

PROOF.[70] It suffices to find a group $\mathfrak{G}$ which fulfills the condition (i) in Lemma 5.3.4 (so that the **M** of §5.3 is a factor, of course, in case (II_1)), (i) in Lemma 5.2.2 (so that **M** is approximately finite by Lemma 5.6.1), and (i) in Lemma 6.1.1 (so that **M** possesses the property Γ). Then Theorem XIV and our remark after Def. 6.1.1 take care of everything.

Now the group $\mathfrak{G} = \tilde{\mathfrak{G}}$ of Lemma 5.6.2 possesses the first two properties. As to (i) in Lemma 6.1.1: Let $a^{(1)}, \cdots, a^{(n)}$ be the given elements of $\mathfrak{G}$. Choose p so that none of the $a^{(1)}, \cdots, a^{(n)}$ move any element other than $1, \cdots, p$.

[70] There exists also a relatively simple direct proof, based on Def. 4.1.1.

Let c_0 be the transposition of $p + 1$ with $p + 2$. Then $c_0 \neq 1$ and it commutes with $a^{(1)}, \cdots, a^{(n)}$.

§6.2 We now proceed to establish the decisive negative result.

LEMMA 6.2.1. *Let a group* $\mathfrak{G}$ *fulfilling* (i) *in Lemma* 5.3.4 *be given, which possesses this property:*[71]

(i) *There exists a set* $\mathfrak{F}_1 \subseteq \mathfrak{G}$ *with these properties:*

(i$_1$) *There exists a* $c_1 \in \mathfrak{G}$ *such that*

$$\mathfrak{F} + c_1\mathfrak{F}c_1^{-1} = \mathfrak{G} - (1).$$

(i$_2$) *There exists a* $c_2 \in \mathfrak{G}$, *such that the three sets* $c_2^l\mathfrak{F}c_2^{-l}$, $l = 0, \pm 1$ *are disjoint.*

Then the **M** *of* §5.3 *does not possess the property* Γ.

PROOF. Assume the opposite, i.e. that **M** possesses the property Γ. Apply Def. 6.1.1 with $n = 2$. Put $A_1 = U_{c_1}$, $A_2 = U_{c_2}$, while $\epsilon > 0$ will be chosen subsequently. Form the $U = U(A_1, A_2; \epsilon)$ described there.

Then we have:

$$Tr_{\mathbf{M}}(U) = 0 \tag{6.2.α}$$

$$[[U^{-1}U_{c_h}U - U_{c_h}]] < \epsilon \qquad \text{for } h = 1, 2. \tag{6.2.β}$$

As U_{c_h}, U are both unitary and

$$U_{c_h}^{-1}U(U^{-1}U_{c_h}U - U_{c_h}) = U - U_{c_h}^{-1}UU_{c_h}.$$

(6.2.β) is equivalent to

$$[[U - U_{c_h}^{-1}UU_{c_h}]] < \epsilon \qquad \text{for } h = 1, 2. \tag{6.2.γ}$$

Now determine the η_c in the sense of Lemma 5.3.2 for U, $U_{c_h}^{-1}UU_{c_h}$, $U - U_{c_h}^{-1}UU_{c_h}$ in succession. If the first is θ_c it is easy to verify that the second is $\theta_{c_hcc_h^{-1}}$[50] and hence the third is $\theta_c - \theta_{c_hcc_h^{-1}}$. Therefore the application of (i) in Lemma 5.3.6 to U and to $U - U_{c_h}^{-1}UU_{c_h}$ gives

$$[[U]]^2 = \sum_{c\in\mathfrak{G}} |\theta_c|^2$$

$$[[U - U_{c_h}^{-1}UU_{c_h}]]^2 = \sum_{c\in\mathfrak{G}} |\theta_c - \theta_{c_hcc_h^{-1}}|^2.$$

As U is unitary, $[[U]]^2 = 1$. Considering this and (6.2.γ) these equations yield

$$\sum_{c\in\mathfrak{G}} |\theta_c|^2 = 1 \tag{6.2.δ}$$

$$\left(\sum_{c\in\mathfrak{G}} |\theta_c - \theta_{c_hcc_h^{-1}}|^2\right)^{\frac{1}{2}} < \epsilon \qquad \text{for } h = 1, 2. \tag{6.2.ϵ}$$

After these preparations we introduce a measure in $\mathfrak{G}$ by defining

$$\nu(\mathfrak{A}) = \sum_{c\in\mathfrak{A}} |\theta_c|^2 \qquad \text{for } \mathfrak{A} \subseteq \mathfrak{G}.$$

[71] This proof copies to a certain extent Hausdorf's famous $1/2 - 1/3$ division of the sphere. In this connection the use of the free group in Lemma 6.2.2 should be noted.

Then (6.2.δ) becomes

$$(6.2.\zeta) \qquad \nu(\mathfrak{G}) = 1$$

(6.2.α) means $\theta_1 = 0$ i.e.

$$(6.2.\eta) \qquad \nu((1)) = 0.$$

The triangle inequality in infinitely many dimensions gives

$$\left| \left(\sum_{c \epsilon \mathfrak{A}} |\theta_c|^2\right)^{\frac{1}{2}} - \left(\sum_{c \epsilon \mathfrak{A}} |\theta_{c_h c c_h^{-1}}|^2\right)^{\frac{1}{2}} \right| \leqq \left(\sum_{c \epsilon \mathfrak{G}} |\theta_c - \theta_{c_h c c_h^{-1}}|^2\right)^{\frac{1}{2}}.$$

The left-hand side is clearly $| \nu(\mathfrak{A})^{\frac{1}{2}} - \nu(c_h \mathfrak{A} c_h^{-1})^{\frac{1}{2}} |$. The right-hand side is $\leqq (\sum_{c \epsilon \mathfrak{G}} |\theta_c - \theta_{c_h c c_h^{-1}}|^2)^{\frac{1}{2}}$ which is $< \epsilon$ by (6.2.ϵ). So we have

$$(6.2.\theta) \qquad | \nu(\mathfrak{A})^{\frac{1}{2}} - \nu(c_h \mathfrak{A} c_h^{-1})^{\frac{1}{2}} | < \epsilon.$$

Now by (6.2.ζ), $\nu(\mathfrak{A})$ and $\nu(c_h \mathfrak{A} c_h^{-1})$ are $\leqq \nu(\mathfrak{G}) = 1$. Hence

$$\begin{aligned} | \nu(\mathfrak{A}) - \nu(c_h \mathfrak{A} c_h^{-1}) | &= | \nu(\mathfrak{A})^{\frac{1}{2}} - \nu(c_h \mathfrak{A} c_h^{-1})^{\frac{1}{2}} | (\nu(\mathfrak{A})^{\frac{1}{2}} + \nu(c_h \mathfrak{A} c_h^{-1})^{\frac{1}{2}}) \\ &\leqq 2 | \nu(\mathfrak{A})^{\frac{1}{2}} - \nu(c_h \mathfrak{A} c_h^{-1})^{\frac{1}{2}} | \end{aligned}$$

Therefore (6.2.θ) becomes

$$(6.2.\iota) \qquad | \nu(\mathfrak{A}) - \nu(c_h \mathfrak{A} c_h^{-1}) | < 2\epsilon.$$

Let us apply (6.2.ι) to $\mathfrak{F}, c_1$, $\mathfrak{F}, c_2$, $c_2^{-1}\mathfrak{F}c_2$, c_2 in place of its $\mathfrak{A}, c_h$. Then

$$(6.2.\kappa) \qquad | \nu(\mathfrak{F}) - \nu(c_1 \mathfrak{F} c_1^{-1}) | < 2\epsilon$$

$$(6.2.\lambda) \qquad | \nu(\mathfrak{F}) - \nu(c_2 \mathfrak{F} c_2^{-1}) | < 2\epsilon$$

$$(6.2.\mu) \qquad | \nu(c_2^{-1} \mathfrak{F} c_2) - \nu(\mathfrak{F}) | < 2\epsilon$$

obtain. Now (i_1) and (6.2.ζ), (6.2.η) and (6.2.κ) give

$$\nu(\mathfrak{F}) + (\nu(\mathfrak{F}) + 2\epsilon) > 1$$

i.e.

$$(6.2.\nu) \qquad \nu(\mathfrak{F}) > \tfrac{1}{2} - \epsilon.$$

On the other hand (i_2) and (6.2.ζ), (6.2.λ) and (6.2.μ) give

$$\nu(\mathfrak{F}) + (\nu(\mathfrak{F}) - 2\epsilon) + (\nu(\mathfrak{F}) - 2\epsilon) < 1$$

i.e.

$$(6.2.o) \qquad \nu(\mathfrak{F}) < \tfrac{1}{3} + \tfrac{4}{3}\epsilon.$$

(6.2.ν) and (6.2.o) imply

$$\tfrac{1}{2} - \epsilon < \tfrac{1}{3} + \tfrac{4}{3}\epsilon \quad \text{or} \quad 1/14 < \epsilon.$$

Hence it suffices to choose $\epsilon = 1/14$ in order to have a contradiction.

Thus we have shown that $\mathbf{M}$ cannot possess the property Γ.

COROLLARY 1. *Under the above assumptions,* **M** *is not approximately finite.*

PROOF. Combine the above lemma with Lemma 6.1.2.

COROLLARY 2. *We may replace condition* (i_2) *on* $\mathfrak{G}$ *and* $\mathfrak{F}$ *by*

(i_2'). *There exist two elements* c_2 *and* c_3 *in* $\mathfrak{G}$ *such that* $\mathfrak{F}$, $c_2\mathfrak{F}c_2^{-1}$, $c_3\mathfrak{F}c_3^{-1}$ *are mutually exclusive.*

PROOF. In the first paragraph of the above proof we let $A_3 = U_{c_3}$ and then take $U = U(A_1, A_2, A_3, \epsilon)$ instead of $U = U(A_1, A_2, \epsilon)$. We have $h = 1, 2, 3$ in (6.2.β), (6.2.γ), (6.2.ϵ), (6.2.θ) and (6.2.ι), instead of $h = 1, 2$. Finally we obtain an equation

$$(6.2.\mu') \qquad | \nu(\mathfrak{F}) - \nu(c_3\mathfrak{F}c_3^{-1}) | < 2\epsilon$$

instead of (6.2.μ), by applying (6.2.ι) to $\mathfrak{F}$, c_3 instead of $c_2^{-1}\mathfrak{F}c_2$, c_2. (6.2.μ') is used instead of (6.2.μ) in order to obtain (6.2.o), and the rest of the proof is the same.

We conclude by giving a specific instance of an **M** which does not possess the property Γ and is not approximately finite. We do this by giving an example of a group $\mathfrak{G}$ which fulfills the requirements of Lemma 6.2.1.

LEMMA 6.2.2. *Let* $\mathfrak{G}$ *be the free group of two generators* c_1, c_2. *Then* $\mathfrak{G}$ *fulfills the requirements of Lemma* 6.2.1.

PROOF. Ad (i) in Lemma 5.3.4: Consider an $a \in \mathfrak{G}$, $a \neq 1$. Write a as a power product of c_1, c_2 of minimum length. Then it is easy to verify that $c_1^{-n}ac_1^{n}$, $n = 1, 2, \cdots$ are pairwise different unless $a = c_1^{m}$. In the latter case we have that the $c_2^{-n}ac_2^{n} = c_2^{-n}c_1^{m}c_2^{n}$ are distinct unless $m = 0$ i.e. $a = 1$. Thus unless $a = 1$, $\mathfrak{K}_a$ is infinite.

Ad (i) in Lemma 6.2.1: Let $\mathfrak{F}$ be the set of $a \in \mathfrak{G}$ which when written as a power product of c_1, c_2 of minimum length end with a c_1^{n}, $n = \pm 1, \pm 2, \cdots$. Then (i_1) and (i_2) in Lemma 6.2.1, i.e. (i) eod., are immediately verified.

Combining Lemma 6.2.2 with Corollary 1 to Lemma 6.2.1 gives

THEOREM XVI. *There exist not approximately finite factors* **M** *in the case* (II_1).

Or, considering Theorem XV,

THEOREM XVI′. *There exist in case* (II_1) *more than one algebraical type* $\bar{\mathbf{M}}$.

§**6.3** We wish in this section to present a number of examples of groups $\mathfrak{G}$ for which the corresponding **M** of §5.3 is not approximately finite.

LEMMA 6.3.1. *Let* $\mathfrak{G}$ *be a group which contains two multiplicative subsets* $\mathfrak{F}_1$ *and* $\mathfrak{F}_2$. *Suppose that* $\mathfrak{G}$ *is the free product of* $\mathfrak{F}_1$ *and* $\mathfrak{F}_2$ *and that* $\mathfrak{F}_1$ *contains at least two elements and* $\mathfrak{F}_2$ *contains at least three. Then* $\mathfrak{G}$ *satisfies condition* (i) *of Lemma* 5.3.4, *condition* (i_1) *of Lemma* 6.2.1 *and condition* (i_2') *of Corollary* 2 *of that lemma* (*for the same* $\mathfrak{F}$).

PROOF. Let c_1 be an element of $\mathfrak{F}_1$ which is not 1, and c_2, c_3 elements of $\mathfrak{F}_2$ which are each not 1.

We prove first that (i) of Lemma 5.3.4 holds, i.e. that for $a \in \mathfrak{G}$, $a \neq 1$ the $\mathfrak{K}_a$ is infinite.

Suppose an $a \in \mathfrak{G}$, $a \neq 1$ is given. Since $\mathfrak{G}$ is a free product of $\mathfrak{F}_1$ and $\mathfrak{F}_2$,

each a has a unique factorization $a_1 \cdot a_2 \cdot \cdots \cdot a_p$ into a product in which the factors are alternately in $\mathfrak{F}_1$ and $\mathfrak{F}_2$ and no factor is 1.

We note for later use that if $c \in \mathfrak{F}_i$, then $c^{-1} \in \mathfrak{F}_i$. Suppose $i = 1$. Then c^{-1} will have a factorization $b_1 \cdot b_2 \cdot \cdots \cdot b_r$. If $b_1 \in \mathfrak{F}_2$ we have $1 = cc^{-1} = c \cdot b_1 \cdot b_2 \cdot \cdots \cdot b_r \neq 1$. Thus $b_1 \in \mathfrak{F}_1$. But $1 = cc^{-1} = (cb_1) \cdot b_2 \cdot \cdots \cdot b_r$ implies $r = 1$. Consequently $c^{-1} = b_1 \in \mathfrak{F}_1$.

Now suppose firstly that a_1 (the first factor) and a_p (the last factor) are both from $\mathfrak{F}_1$. Let $b_n = (c_2c_1)^n$. Then the set of $b_n^{-1}ab_n$ are for $n = 1, 2, \cdots$ all distinct. (The factorization is immediately apparent.) Hence $\mathfrak{K}_a$ is infinite in this case.

If a_1 and a_p are both from $\mathfrak{F}_2$, we let $b_n = (c_1c_2)^n$ and the result is again apparent.

If a_1 is from $\mathfrak{F}_1$ and a_p is from $\mathfrak{F}_2$, we consider c_2 and c_3 and denote by c' that one of c_2, c_3 which is not equal to a_p^{-1}. Let $b_n = (c'c_1)^n$. Then for $n = 1, 2, \cdots$ the factorization of $b_n^{-1}ab_n$ is obtained by simply writing out the expressions for these elements and coalescing a_p and the first c' in $b_n.a_pc'$ is in $\mathfrak{F}_2$ but is not 1 since $c' \neq a_p^{-1}$. Hence the factorization has been obtained. These factorizations for $n = 1, 2, \cdots$ show that the $b_n^{-1}ab_n$ are distinct, and hence $\mathfrak{K}_a$ is infinite.

If a_1 is in $\mathfrak{F}_2$ and a_p is in $\mathfrak{F}_1$ we define c' as that one of c_2, c_3 which is not equal to a_1. An argument similar to that of the preceding paragraph will show that in this case also $\mathfrak{K}_a$ is infinite.

Thus we have established (1) of Lemma 5.3.4.

To establish (i_1) of Lemma 6.2.1 and (i_2') of Corollary 2 of that lemma for the same $\mathfrak{G}$, we define $\mathfrak{F}$ as the set of a's, $a \neq 1$ having a factorization $a_1 \cdot \cdots \cdot a_p$ in which a_p is in $\mathfrak{F}_1$.

We show firstly (i_1) in Lemma 6.2.1, i.e.

$$\mathfrak{F} + c_1^{-1}\mathfrak{F}c_1 = \mathfrak{G} - (1).$$

If $a \in \mathfrak{G}$, $a \neq 1$ then a has a factorization $a_1 \cdot \cdots \cdot a_p$. Now if a is not $\in \mathfrak{F}$, a_p is not $\in \mathfrak{F}_1$ but $a_p \in \mathfrak{F}_2$. In this case we consider $c_1ac_1^{-1} = c_1a_1 \cdot \cdots \cdot a_pc_1^{-1}$. If p is even, i.e. $p = 2, 4, \cdots$ then $a_1 \in \mathfrak{F}_1$ (since $a_p \in \mathfrak{F}_2$) and the factorization of $c_1ac_1^{-1}$ is either $(c_1a_1) \cdot a_2 \cdot \cdots \cdot a_p \cdot c_1^{-1}$ or (if $c_1a_1 = 1$) $a_2 \cdot \cdots \cdot a_pc_1^{-1}$. If p is odd, i.e., $p = 1, 3, \cdots$, $a_1 \in \mathfrak{F}_2$ and $c_1ac_1^{-1}$ has the factorization $c_1 \cdot a_1 \cdot \cdots \cdot a_p \cdot c_1^{-1}$. These factorizations all show that $c_1ac_1^{-1}$ is $\in \mathfrak{F}$. Thus if $a \neq 1$ and a is not $\in \mathfrak{F}$, $c_1ac_1^{-1} \in \mathfrak{F}$ or $a \in c_1^{-1}\mathfrak{F}c$. This is equivalent to $\mathfrak{G} - (1) \subseteq \mathfrak{F} + c_1^{-1}\mathfrak{F}c_1$. However this inclusion is not proper. For 1 is not in $\mathfrak{F}$. Consequently 1 is not in $c_1^{-1}\mathfrak{F}c_1$ and $\mathfrak{F} + c_1^{-1}\mathfrak{F}c_1$. We can now conclude

$$\mathfrak{F} + c_1^{-1}\mathfrak{F}c_1 = \mathfrak{G} - (1).$$

One can establish (i_2') of Corollary 2 of Lemma 6.2.1 in a similar (indeed easier) way.

Thus our Lemma has been proven.

This lemma permits us to offer many other examples similar to the $\mathfrak{G}$ of Lemma

6.2.2. For we may obtain $\mathfrak{F}_1$ and $\mathfrak{F}_2$ in a number of ways, and then $\mathfrak{G} = \mathfrak{G}(\mathfrak{F}_1, \mathfrak{F}_2)$ is constructed by adding to 1 the "free" products $a_1 \cdot a_2 \cdot \cdots \cdot a_p$ where the a_i's are never 1 and are alternately from $\mathfrak{F}_1$ and $\mathfrak{F}_2$. As an example we may take $\mathfrak{F}_1$ as the group generated by a single generator a of order $p = 2, 3, \cdots, \infty$ and $\mathfrak{F}_2$ as the group generated by an element b of order $q = 3, 4, \cdots, \infty$. $\mathfrak{G}(a, b) = \mathfrak{G}(\mathfrak{F}_1, \mathfrak{F}_2)$ is then an example. Of course we may take for $\mathfrak{F}_1$ any group of order $\geqq 2$ and $\mathfrak{F}_2$ of order $\geqq 3$.

The restriction that $\mathfrak{F}_2$ contains at least 3 members rules out the case in which $\mathfrak{F}_1$ and $\mathfrak{F}_2$ are both groups of order 2, i.e. $\mathfrak{F}_1$ consists of $(1, a)$ and $\mathfrak{F}_2$ consists of $(1, b)$. Here of course we do not have the (i_2) of Lemma 6.2.1 or the (i_2') of its second corollary. However, condition (i) of Lemma 5.3.4 also fails. For the element ab has only ba and itself as conjugates, i.e. $\mathfrak{K}_{ab} = (ab, ba)$. Thus $\mathfrak{K}_{ab}$ is finite. Hence by Lemma 5.3.4, the **M** for this $\mathfrak{G}$ is not even a factor.

Another lemma, which we shall use in an appendix, is readily shown now.

Lemma 6.3.2. *If $\mathfrak{G}$ is such that to every $a \in \mathfrak{G}$, $a \neq 1$ there is a $b \in \mathfrak{G}$, b of order $\geqq 3$ such that $\mathfrak{G}$ contains the free group $\mathfrak{G}(a, b)$ then $\mathfrak{G}$ fulfills* (i) *of Lemma* 5.3.4. (*If a is not of order 2, b need only differ from the identity.*)

Proof. Let $\bar{\mathfrak{K}}_a$ denote the class of a relative to the subgroup $\mathfrak{G}(a, b)$. Our discussion following Lemma 6.3.1 shows that $\mathfrak{G}(a, b)$ satisfies the hypotheses of Lemma 6.3.1. Consequently that Lemma and Lemma 5.3.4 imply that $\bar{\mathfrak{K}}_a$ is infinite. Since $\bar{\mathfrak{K}}_a \subseteq \mathfrak{K}_a$ we have $\mathfrak{K}_a$ infinite for $\mathfrak{G}$ itself, and thus (i) of Lemma 5.3.4 holds.

Appendix[72]

In this appendix we give examples of two Class (II_1) factors which are not isomorphic and yet each of which is isomorphic to a subset of the other.

The factors will be the factors associated with certain groups $\mathfrak{G}_1$ and $\mathfrak{H}_1$ as in §5.3. These groups are best presented as subgroups of a group $\mathfrak{G}$ which we shall now define.

$\mathfrak{G}$ is determined by generators x_r one for each rational number $r = p/q$. The commutation rules for these generators are determined as follows. Let A denote the set of dyadic rationals $q/2^n$ in the most reduced form where $n = 0, 1, 2, \cdots$ and q is odd except possibly when $n = 0$. Let B denote the set of all other rationals.

The commutation rules for x_r are the following:

(i) If $r \in A$ then x_r combines freely with any product $x_{s_1} \cdot x_{s_2} \cdot \cdots \cdot x_{s_p}$ if $s_1, s_2, \cdots, s_p$ are all $< r$.

(ii) If $r \in B$, x_r commutes with all products $x_{s_1} \cdot x_{s_2} \cdot \cdots \cdot x_{s_p}$ if $s_1, s_2, \cdots, s_p$ are all $< r$.

(iii) All x_r are of order ∞.

One can show that the set of products $x_{r_1}^{n_1} \cdot \cdots \cdot x_{r_p}^{n_p}$, $n_i = \pm 1, \pm 2, \cdots$ (with

[72] This example was originally part of a later investigation by the second-named author, but it can be most easily presented here, using our present methods.

a finite number of factors) constitutes a denumerably infinite group $\mathfrak{G}$.[73] This is also true for the set of products in which the x_r's are restricted, so that r is in a fixed non-empty subset S of the rational numbers.

Let $\mathfrak{H}_r$ be the subgroup of $\mathfrak{G}$ determined by the generators $x_{r'}$ with $r' \leqq r$. Let $\mathfrak{G}_r$ be the subgroup of $\mathfrak{G}$ determined by the generators $x_{r'}$ with $r' < r$. We can show

LEMMA A.1. (i) $\mathfrak{G}_r \subsetneqq \mathfrak{H}_r$

(ii) *If* $r' < r$, $\mathfrak{H}_{r'} \subsetneqq \mathfrak{G}_r$.

(iii) *If* $n = \pm 1, \pm 2, \cdots$ *then* $\mathfrak{H}_{r+n}$ *is isomorphic to* $\mathfrak{H}_r$.

(iv) $\mathfrak{G}_r$ *has property* (i) *of Lemma* 5.3.4.

(v) $\mathfrak{G}_r$ *has property* (i) *of Lemma* 6.1.1.

(vi) *Assume* $r \in A$. *Then* $\mathfrak{H}_r$ *satisfies the hypotheses of Lemma* 6.3.1, *and consequently it satisfies* (i_1) *and* (i_2') *of Lemma* 6.2.1 *and its Corollary* 2 *and* (i) *in Lemma* 5.3.4.

PROOF. Ad (i) and (ii) are clear.

Ad (iii). The correspondence $x_{r+n} \sim x_r$ between the generators of $\mathfrak{H}_{n+r}$ and $\mathfrak{H}_r$ is one-to-one and preserves the properties $r \in A$ and $r \in B$. Consequently it determines an isomorphism between $\mathfrak{H}_{n+r}$ and $\mathfrak{H}_r$.

Ad (iv). Suppose $a \in \mathfrak{G}_r$. Then $a = x_1^{n_1} \cdot \cdots \cdot x_{r_p}^{n_p}$ with each $r_i < r$. Let r' be the $\mathrm{Max}_{i=1,\cdots,p}\, r_i$. Then of course $r' < r$ and there is an $r'' \in A$ such that $r' < r'' < r$. By the first commutation rule above $x_{r''}$ combines freely with a. By the definition of $\mathfrak{G}_r$, $\mathfrak{G}(a, x_{r''}) \subsetneqq \mathfrak{G}_r$. The third rule above tells us that x_r is of infinite order. These last two results show that the hypotheses of Lemma 6.3.2 are satisfied. It follows from that lemma that the condition (i) of Lemma 5.3.4 is satisfied.

Ad (v). $\mathfrak{G}_r$ has property (i) of Lemma 6.1.1: For let $a_1, \cdots, a_k$ be any k elements of $\mathfrak{G}_r$ and let $x_{r_1}, x_{r_2}, \cdots, x_{r_l}$ denote the generators which appear explicitly in $a_1, \cdots, a_k$. Let $r' = \mathrm{Max}_{i=1,\cdots,l}\, r_i$. Then $r' < r$ by the definition of $\mathfrak{G}_r$. Hence there is an $r'' \in B$ such that $r' < r'' < r$. Let $c = x_{r''}$. Then $c \neq 1$, $c \in \mathfrak{G}_r$ and c commutes with $a_1, \cdots, a_k$ by the second commutation rule above. This result shows that (i) in Lemma 6.1.1 is satisfied.

Ad (vi). For $\mathfrak{H}_r$ we let $\mathfrak{F}_1 = \mathfrak{G}_r$ and $\mathfrak{F}_2 = (x_r^n)$. Now x_r is of infinite order and the first commutation rule shows that $\mathfrak{H}_r$ is the free product of $\mathfrak{G}_r$ and $\mathfrak{F}_2$. These show that the hypotheses of Lemma 6.3.1 are fulfilled. That lemma itself

---

[73] This can be most readily shown by obtaining a "normal form" for each such product $x_{r_1}^{n_1} \cdot \cdots \cdot x_{r_p}^{n_p}$. This normal form may be obtained as follows. Let r_i denote the largest subscript which appears and is in B. We move $x_{r_i}^{n_i}$ as far to the left as possible. Let $r_{i'}$ be the next largest subscript, which is also in B. We then move $x_{r_{i'}}^{n_{i'}}$ as far to the left as possible. We continue on until we have considered all the $r_i \in B$. The resulting product is then in a form in which the $x_{r_j}^{n_j}$, $r_j \in A$ have the same order as before, but each such $x_{r_j}^{n_j}$, $r_j \in A$ is followed by a product $x_{r_{i_1}}^{n_{i_1}} \cdot \cdots \cdot x_{r_{i_k}}^{n_{i_k}}$ with the r_{i_t} each in B, each $< r_j$ and with $r_{i_1} < r_{i_2} < \cdots < r_{i_k}$.

We may use this form to establish the associativity of the composition rule. The existence of an inverse is clear.

now shows that (i_1) and (i_2') of Lemma 6.2.1 and its Corollary 2 are satisfied, and (i) in Lemma 5.3.4.

We next state a quite general lemma.

LEMMA A.2. *Let* $\mathbf{M}$ *be a ring which contains* $(\alpha 1)$ *and suppose* $f \in \mathfrak{H}$ *is such that* $Af = 0$ *and* $A \in \mathbf{M}$ *imply* $A = 0$. *Let* $\mathfrak{M} = \mathfrak{M}_f^{\mathbf{M}}$. *Then if* A *is closed, has domain dense, and* $\eta\ \mathbf{M}$, A *is bounded if and only if it is bounded on* $\mathfrak{M}$.

PROOF. It is clear that we need only show that if A is unbounded, it is unbounded on $\mathfrak{M}$. Let $A = WB$ where W is partially isometric and B is definite (cf. [5], §4.4). Since $A\ \eta\ \mathbf{M}$, we have $B\ \eta\ \mathbf{M}$ and also for all g, $|Ag| = |Bg|$. (This may be ∞ of course.) Hence we may substitute B for A, or what is the same thing assume that A is definite. But A definite implies that there is a resolution $\{E(\lambda)\} \subset \mathbf{M}$ such that $A = \int_0^\infty \lambda\, dE(\lambda)$. Since A is unbounded, $1 - E(\lambda) \neq 0$ for $\lambda < \infty$. Let $g = (1 - E(\lambda))f$ for any fixed λ. By our hypothesis on f, $g \neq 0$ since $1 - E(\lambda) \neq 0$. From the formula for A we see $|Ag| > \lambda |g|$. Furthermore $g \in \mathfrak{M}_f^{\mathbf{M}} = \mathfrak{M}$. Thus if A is unbounded, it is unbounded on $\mathfrak{M}$.

COROLLARY. *If* $A \in \mathbf{M}$ *the bound of* A *on* $\mathfrak{M}$ *is the same as the bound of* A ($|||A|||$).

PROOF. It is clear that the bound of A on $\mathfrak{M}$ is $\leqq |||A|||$. To show that it can't be less one uses an argument similar to the above.

We wish to apply Lemma A.2 to the case in which $\mathbf{M}$ is a subring of the set of group numbers in the sense of §5.3. We may do this by taking $f = \varphi_1$ in the sense of (5.3.α). For if $A \in \mathbf{M}$, we have that $A \sim \{\eta_{a^{-1}b}\}$ by Lemma 5.3.2 and $A\varphi_1 = \sum_{c \in \mathfrak{G}} \eta_c \varphi_c$ by (^{49}, γ). Thus $A\varphi_1 = 0$ implies $\eta_c = 0$ for all $c \in \mathfrak{G}$ and hence $A = 0$.

It is also useful to notice in applying this to group numbers that if we have any sequence $\{\eta_c\}$ with $\sum_{c \in \mathfrak{G}} |\eta_c|^2 < \infty$ then the matrix $\{\eta_{a^{-1}b}\}$ determines an A with a closure and domain dense. For it is clear from (^{49}, γ) that A is defined for all φ_a, $a \in \mathfrak{G}$ of (5.3.α) and hence for the linear combinations of these. Since the latter are dense, A has domain dense. Furthermore the matrix $\{\bar{\eta}_{b^{-1}a}\}$ is adjoint to it. Since this latter also has domain dense, A must have a closure. We shall consider A to be the least closed transformation, with matrix $\{\eta_{a^{-1}b}\}$. On its domain, A is the limit of finite linear combinations of the U_c and hence is $\eta\ \mathbf{M}$. Cf. Lemmas 5.3.1, 5.3.2 and 5.3.6.

LEMMA A.3. *Suppose* $\mathfrak{G}^0$ *and* $\mathfrak{G}^{00}$ *are two groups and* $\mathbf{M}^0$ *and* $\mathbf{M}^{00}$ *their respective group numbers in the sense of* §5.3. *Then if* $\mathfrak{G}^0 \subsetneqq \mathfrak{G}^{00}$, $\mathbf{M}^0$ *is algebraically isomorphic with a subring of* $\mathbf{M}^{00}$.

PROOF. For an $A \in \mathbf{M}^{00}$ we have a sequence $\{\eta_c\,;\, c \in \mathfrak{G}^{00}\}$ as in Lemma 5.3.2. Let $\mathbf{M}^{0+}$ denote the set of A's $\in \mathbf{M}^{0,0}$ for which $\eta_c = 0$ if c not $\in \mathfrak{G}^0$. Since $\mathfrak{G}^0$ is a subgroup, this set of A's is closed under the operations αA, A^*, $A + B$, AB. Furthermore this set of A's is metrically closed, and as we mentioned at the end of §5.5, for a subalgebra of group numbers, this implies ring closure. Thus $\mathbf{M}^{0+}$ is a ring, and indeed is the ring determined by the U_c, $c \in \mathfrak{G}^0$.

Let $\mathfrak{M}^0$ be the set $\subseteq \mathfrak{H}^{0,0}$ determined by the φ_a with $a \in \mathfrak{G}^0$ (cf. 5.3.α). It is clear that we can identify this with $\mathfrak{H}^0$ the space on which we have $\mathbf{M}^0$ defined. Now for every $A \in \mathbf{M}^0$ we have a sequence $\{\zeta_c ; c \in \mathfrak{G}^0\}$ and corresponding matrix $\{\zeta_{a^{-1}b}\}$. Correspondingly, we have a sequence $\{\eta_c , c \in \mathfrak{G}^{00}\}$ with $\eta_c = \zeta_c$ for $c \in \mathfrak{G}^0$ and $\eta_c = 0$ otherwise. As we pointed out above, the matrix $\{\eta_{a^{-1}b}\}$ determines a closed operator $\tilde{A}$ with domain dense. This is clearly $\eta\, \mathbf{M}^{0+}$ and hence by the application of Lemma A.2 described above, it is bounded if and only if it is bounded on $\mathfrak{M}_f^{\mathbf{M}^{0+}}$. Now one can show by our above definition of $\mathbf{M}^{0+}$ that $\mathfrak{M}_f^{\mathbf{M}^{0+}} = \mathfrak{M}^0$. Now $\tilde{A} = A$ on $\mathfrak{M}^0$ and thus is bounded there. Hence, by Lemma A.2, $\tilde{A}$ is bounded and $\in \mathbf{M}^{0+}$. We now see that to every $A \in \mathbf{M}^0$ there corresponds an $\tilde{A} \in \mathbf{M}^{0+}$ such that $\tilde{A}$ agrees with A on $\mathfrak{H}^0 = \mathfrak{M}^0$. On the other hand, it is clear that to every $\tilde{A} \in \mathbf{M}^{0+}$ we have an $A \in \mathbf{M}^0$ with this property.

One can readily verify that this is an algebraic isomorphism.

We are now ready for the final step. Let $\mathbf{M}_1$ be the set of group numbers for $\mathfrak{G}_1$, $\mathbf{M}_2$ the set of group numbers for $\mathfrak{H}_1$ (cf. §5.3).

LEMMA A.4. (i) $\mathbf{M}_1$ *and* $\mathbf{M}_2$ *are factors in Case* (II_1).

(ii) $\mathbf{M}_1$ *is algebraically isomorphic with a subring of* $\mathbf{M}_2$.

(iii) $\mathbf{M}_2$ *is algebraically isomorphic with a subring of* $\mathbf{M}_1$.

(iv) $\mathbf{M}_1$ *has property* Γ *of Def.* 6.1.1.

(v) $\mathbf{M}_2$ *does not have property* Γ.

PROOF. Ad (i). $\mathbf{M}_1$ is a factor by Lemma A.1 (iv) and Lemma 5.3.4. $\mathbf{M}_2$ is a factor by Lemma A.1 (vi) and Lemma 5.3.4. $\mathbf{M}_1$ and $\mathbf{M}_2$ are both in Case (II_1) by Lemma 5.3.5 (i).

Ad (ii). This follows from Lemma A.1 (i) and Lemma A.3.

Ad (iii). Let $\mathbf{M}_0$ denote the set of group numbers for $\mathfrak{H}_0$. Lemma A.1 (iii) for $n = -1$ shows that $\mathfrak{H}_0$ and $\mathfrak{H}_1$ are isomorphic, and this of course implies that $\mathbf{M}_2$ is algebraically isomorphic to $\mathbf{M}_0$. Lemma A.1 (ii) and Lemma A.3 imply that $\mathbf{M}_0$ is isomorphic to a subring of $\mathbf{M}_1$. Hence $\mathbf{M}_2$ is algebraically isomorphic to a subring of $\mathbf{M}_1$.

Ad (iv). This is implied by Lemma A.1 (v) and Lemma 6.1.1.

Ad (v). This is implied by Lemma A.1 (vi) and Corollary 2 of Lemma 6.2.1.

Since property Γ is an isomorphism invariant, (iv) and (v) imply that $\mathbf{M}_1$ and $\mathbf{M}_2$ are not algebraically isomorphic. Combining this with (ii), (i) and (iii), we have

THEOREM A. *There exist two rings* $\mathbf{M}_1$, $\mathbf{M}_2$, *each a factor in Case* (II_1), *which are not algebraically isomorphic to each other but are such that each is algebraically isomorphic to a subring of the other.*

THE INSTITUTE FOR ADVANCED STUDY AND
COLUMBIA UNIVERSITY

INDEX

INDEX

ERRATUM

Page 318, lines 12 and 13 (after semicolon). Read: "*and therefore, if B is completely additive* (*i. e., if the union of the elements of every subset of B exists*), *then it is isomorphic with a direct sum of f-homogeneous Boolean algebras.*"